

文章编号:1000-4653(2025)04-0121-11

基于二次分解和算法优化 LSTM 的 港口集装箱吞吐量预测研究

王博文, 黄逸, 孟铨博, 曹添悦

(交通运输部东海航海保障中心上海海事测绘中心, 上海 200090)

摘要:准确预测港口集装箱吞吐量对港口航运企业和政府管理部门科学制定决策具有重要意义。目前的研究方法对短历时港口集装箱吞吐量的关注较少,对非线性、非平稳的波动序列的预测准确性有限。本文以上海港集装箱吞吐量为分析与预测对象,研究提出了基于以相关系数分析为基础的变分模态分解(CCVMD)和季节趋势分解(STL)的二次分解的新型深度学习模型:以相关系数为参照,对原始时间序列进行变分模态分解,在此基础上二次分解为季节项、趋势项和不规则项,并用算法优化的长短期记忆神经网络分别对分解项进行预测,汇总得到最终预测结果。结果表明:在集装箱吞吐量预测中,对数据进行预处理的分解组合模型表现显著优于其他模型;本文提出的模型的平均绝对百分比误差为0.021 703,均方根误差百分比为0.026 852,平均绝对误差百分比为0.022 14,预测整体表现优于其余12种比较模型和既往研究提出的部分模型;二次分解预处理在追踪极值、除噪降噪和可解释性方面更具可靠性。

关键词:集装箱吞吐量预测;二次分解;CCVMD;STL;HHO;LSTM;深度学习

中图分类号:U692 文献标志码:A DOI:10.3969/j.issn.1000-4653.2025.04.014

Research on port container throughput forecasting based on secondary decomposition and LSTM optimized by algorithm

WANG Bowen, HUANG Yi, MENG Xuanbo, CAO Tianyue

(Shanghai Maritime Surveying and Mapping Center, Eastern Navigation Service Center, Maritime Safety Administration, Shanghai 200090, China)

Abstract: Accurate forecasting of port container throughput is of great significance for port operators and government administrations in making scientific decisions. Existing forecasting methods, however, often pay insufficient attention to short-calendar-time PCT and exhibit limited accuracy in handling nonlinear and non-stationary fluctuation series. This paper takes the container throughput of Shanghai Port as the research object and proposes a novel deep learning model based on secondary decomposition using CCVMD and STL. Using the correlation coefficient as a reference, variational mode decomposition is first applied to the original time series. Subsequently, a secondary decomposition divides the data into seasonal, trend, and residual components. An algorithm-optimized long short-term memory neural network is then employed to predict each component separately, and the final prediction results are aggregated. Experimental results show that the combined decomposition model with data preprocessing significantly outperforms other models in PCT forecasting. The proposed model achieves a mean absolute percentage error of 0.021 703, a root mean square error percentage of 0.026 852, and a mean absolute error percentage of 0.022 14, indicating superior overall performance compared to 12 benchmark models and several models from prior studies. Furthermore, the secondary decomposition approach demonstrates enhanced

收稿日期:2024-12-21

通信作者:王博文(1989—),女,工程师,研究方向为航海保障。E-mail:wbw_sh@sina.com

引用格式:王博文,黄逸,孟铨博,等.基于二次分解和算法优化 LSTM 的港口集装箱吞吐量预测研究[J].中国航海,2025,48(4):121-131.

WANG B W, HUANG Y, MENG X B, et al. Research on port container throughput forecasting based on secondary decomposition and LSTM optimized by algorithm[J]. Navigation of China, 2025, 48(4): 121-131. (in Chinese)

reliability in tracking extreme values, removing and reducing noise, and improving interpretability.

Key words: container throughput forecasting; secondary decomposition; CCVMD; STL; HHO; LSTM; deep learning

海运具有运能大、成本低、污染小、占地少等优势,是当前国际贸易往来的重要运输形式。全球超过80%的商品贸易通过海上运输,其中,集装箱运输占总运输量的35%左右,占商业价值的60%以上,而且所占比例越来越大^[1]。港口的吞吐量水平还影响到港口技术效率和服务水平^[2],从而对港口竞争力产生影响。合理预测国家或地区重要港口的集装箱吞吐量,准确掌握港口物流需求,有助于港口航运企业合理调整集装箱船运力,科学配置生产要素,也有助于管理部门稳妥制定和调整产业政策、贸易政策,在国际贸易竞争中占据先发优势。

然而,港口集装箱吞吐量受到国家出口贸易量^[3]、可用土地供应量^[4]、国际政治环境等宏观因素,船舶周转时间、船舶吃水、集装箱停留时间、泊位生产率、集装箱存储能力和定制申报时间^[5]等微观因素,以及COVID-19等突发事件的影响,具有非线性、非平稳的特点^[6]。因此,及时准确地预测港口集装箱吞吐量存在较大挑战。此外,大量的实证研究表明,以港口集装箱吞吐量年度数据为分析对象的长期预测往往表现出相对较好的准确性,而以月度数据为分析对象的短期预测甚至年度预测,则存在更大的随机性和波动性。因而,如何提升短历时(预测期1年左右)港口集装箱吞吐量预测可靠性,以满足港口及航运有关方需求,也是该研究领域值得关注的问题。

鉴于此,本文构建了以相关系数分析为基础的变分模态分解(Correlation Coefficient-Variational Mode Decomposition, CCVMD)和季节趋势分解(Seasonal and Trend decomposition using Loess, STL)相结合的二次分解模式,使用长短期记忆神经网络(Long Short-Term Memory, LSTM)预测分解序列,并用 Harris Hawk Optimization, HHO)算法确定 LSTM 的隐含层神经元数量、最大训练次数和初始学习率等超参数的最优解,以减少主观判断偏差对模型预测精度的干扰。将所提出的模型与7种单一模型和5种组合模型,以及既有研究中的部分推荐模型进行预测效果比较,结果表明所提出的模型在评价指标上展现出更好的预测性能。

1 文献综述

港口集装箱吞吐量预测的方法从整体上可分为定性预测方法和定量预测方法。

1.1 定性预测

定性预测方法依靠专家学者、从业人员等的经验知识和主观判断,对分析对象未来的发展趋势进行预测,常见的方法包括调查判断预测法、德尔菲法和交叉影响分析预测法等。定性预测方法对数据资料的依赖少,易于理解和操作,灵活性高,但其预测结果受个人经验与偏好的影响显著,在预测问题上存在较大的局限性^[7]。

1.2 定量预测

定量预测方法根据使用的模型种数,分为单一模型预测与组合模型预测。

单一模型预测一次只采用一种模型进行预测分析。支持向量回归(Support Vector Regression, SVR)、LSTM、季节性自回归综合移动平均(Seasonal Autoregressive Integrated Moving Average, SARIMA)等模型^[8],移动平均法、多元自适应回归样条、自回归综合移动平均法和简单指数平滑等时间序列预测方法^[9-10],是较为常见成熟的预测模式。由于单一模型往往难以准确捕捉经济社会活动中复杂数据间的内在关系,存在欠拟合风险较高、鲁棒性较差等问题。组合预测可以减少预测误差,并降低模型的不确定性^[11],并通过组合实现了单个预测多样性的增加^[12]。

组合模型预测按照是否对数据进行预处理,分为简单组合模型预测和分解组合模型预测。简单组合模型预测通常由不同的预测模型或算法与预测模型组合而成,如以自回归综合移动平均(Auto Regressive Integrated Moving Average, ARIMA)模型为基础简单加权组合和残差优化组合^[13],随机森林算法和双向 LSTM 组合^[14],基于主成分分析和多元线性回归的组合预测^[15]。虽然将多个模型组合在一起,有助于发挥每个模型的优势,提高模型的拟合性和鲁棒性,但随着时间序列样本数据的增加,样本的波动性和不确定性增大,简单组合模型很难从大量复杂数据中准确提炼能够反映样本特征的重要信息。分解组合模型预测在简单组合模型的基础上,考虑了对时间序列的预处理,运用分解算法削弱时间序列中的噪声影响,从而进一步提高了模型预测性能。以变分模态分解(Variational Mode Decomposition, VMD)为基础的、由蝴蝶优化算法优化的极限学习机^[16],融合 VMD、SARIMA、卷积神经网络、LSTM 和注意力机制等方法的分解组合模型^[17],基于经验模态分解和选择性深度集成的组合模型^[18],

经验小波变换分解和时空卷积网络的组合^[19],时间层次结构、横截面层次结构、跨时空层次结构等概念分解^[20],以及粒子群优化 VMD^[21]等分解组合模型是当下集装箱吞吐量预测研究的热门方向。

1.3 当前研究不足

根据文献综述,目前的研究领域还存在如下问题:大多数研究侧重于长期预测,对短期特别是1年之内的短历时预测关注不多;基于二次分解的组合模型在港口集装箱吞吐量预测的研究中应用较少;VMD 参数选取或依赖于经验判断,或虽采用优化算法择定超参数,但未对组合模型中使用的机器学习方法进行算法优化,限制了预测效果提升;分解模态序列的选取也往往依赖于基于图像的主观判断,不够客观可靠。

鉴于此,本文提出了基于 CCVMD-STL 二次分解的新型深度学习模型:以上海港连续 12 个月的短期港口集装箱吞吐量为预测对象,以相关系数为准则对 VMD 分解模态序列进行筛选、组合,经 STL 二次分解后导入 HHO 优化的 LSTM 实现预测。

2 方法论

本文提出的二次分解深度学习模型由 CCVMD、STL、HHO 和 LSTM 组成。主要研究步骤如下。

1) 将上海港集装箱吞吐量月度数据输入 VMD,根据中心频率确定分解模态数,得到各模态分量。

2) 计算各分解模态序列与集装箱吞吐量时间序列之间的相关系数,选取相关系数大于给定阈值的分解模态并重组,得到去噪后新的时间序列。

3) 运用 STL 将去噪时间序列分解成趋势项、季节项和不规则项。

4) 用 HHO 算法优化 LSTM 的隐含层神经元数量、最大训练次数和初始学习率等超参数,并分别预测 STL 分解项。

5) 将 LSTM 预测结果重组,得到集装箱吞吐量最终预测值。

$$X(t+1) = \begin{cases} X_{\text{rand}}(t) - r_1 |X_{\text{rand}}(t) - 2r_2 X(t)|, q \geq 0.5 \\ (X_{\text{rabbit}}(t) - X_m(t)) - r_3(lb + r_4(ub - lb)), q < 0.5 \end{cases} \quad (2)$$

开发阶段,个体位置更新模式如式(3)。

$$X(t+1) = \begin{cases} X_{\text{rabbit}}(t) - X(t) - E |2(1 - r_5)X_{\text{rabbit}}(t) - X(t)|, |E| \geq 0.5, r \geq 0.5 \\ X_{\text{rabbit}}(t) - E |X_{\text{rabbit}}(t) - X(t)|, |E| < 0.5, r \geq 0.5 \\ Y, f(Y) \leq f(Z), r < 0.5 \\ Z, f(Y) > f(Z), r < 0.5 \end{cases} \quad (3)$$

模型组成部分的相关理论介绍如下。

2.1 CCVMD

VMD 是 Konstantin Dragomiretskiy 提出的一种自适应、非递归的模态变分和信号处理方法。该方法在迭代搜索过程中,可以自适应地确定每个模态分量的最佳中心频率和带宽,克服了经验模态分解(Empirical Mode Decomposition, EMD)存在的端点效应、模态混叠等问题,具有更好的可解释性和鲁棒性。

SHEN 等^[22]在全球定位系统和重力恢复与气候实验的研究中,引入相关系数以提取 VMD 分解得到的特征模态分量。CCVMD 是对经 VMD 处理得到的模态分量(Intrinsic Mode Function, F_{IM_i}),引入模态分量序列和原始时间序列的相关系数:

$$R_i = \frac{\text{Cov}(S_T, F_{\text{IM}_i})}{\sqrt{\text{Var}(S_T) \times \text{Var}(F_{\text{IM}_i})}} \quad (1)$$

式(1)中, S_T 为原始时间序列; R_i 为第 i 个 F_{IM_i} 与 S_T 的相关系数; Cov 为协方差; Var 为方差。相关系数越大,则模态分量序列对原始时间序列的解释性越强,关联度也越大。

2.2 STL

STL 是一种稳健的时间序列分解方法,由 Cleveland R B 等人于 1990 年提出。STL 运用局部多项式回归拟合方法提取时间序列中的趋势项、季节项和不规则项,以便于更好地挖掘数据结构特点和内在规律。

2.3 HHO 算法

HHO 算法是由 Heidari A A 等人于 2019 年提出的仿生优化算法,相较于既往不少研究已经涉及的遗传算法、粒子群算法、灰狼算法等,具有全局搜索能力较强、更易跳出局部最优、收敛速度较快、表现相对稳定等优点。该算法模拟哈里斯鹰群体狩猎过程,包括搜索阶段、转换阶段和开发阶段。在转换阶段,哈里斯鹰根据猎物逃跑能量判断进入搜索或者开发阶段。

搜索阶段,个体位置更新模式如式(2)。

$$Y = \begin{cases} X_{\text{rabbit}}(t) - E |2(1 - r_5)X_{\text{rabbit}}(t) - X(t)|, & |E| \geq 0.5, r < 0.5 \\ X_{\text{rabbit}}(t) - E |2(1 - r_5)X_{\text{rabbit}}(t) - X_m(t)|, & |E| < 0.5, r < 0.5 \end{cases} \quad (4)$$

$$Z = Y + S \times LF(D) \quad (5)$$

式(2)~(6)中, $X(t)$ 、 $X_{\text{rabbit}}(t)$ 、 $X_m(t)$ 、 $X_{\text{rand}}(t)$ 分别为第 t 次迭代时的哈里斯鹰个体位置、拥有最优适应度的个体位置、群体平均位置和随机选择的个体位置; Y 、 Z 为第 $t+1$ 次迭代时的哈里斯鹰个体位置判断条件; r 、 r_1 、 r_2 、 r_3 、 r_4 、 r_5 和 q 为取值范围 $[0, 1]$ 的随机数; ub 、 lb 分别为搜索范围的上界和下界; E 为猎物逃跑能量; f 为适应度函数; S 为 $1 \times D$ 的随机向量; $LF(D)$ 为与个体位置的维度相关的飞行函数。

2.4 LSTM

LSTM 是一种特殊的循环神经网络 (Recurrent Neural Network, RNN), 相较于传统 RNN, LSTM 可以有效传递长时间序列中的信息, 不会因为时间序列过长而导致有效信息遗忘, 并且能较好地解决梯度消失和梯度爆炸问题。LSTM 通过遗忘门、输入门和输出门控制信息的提取、传递和输出, 基本结构如图 1 所示。

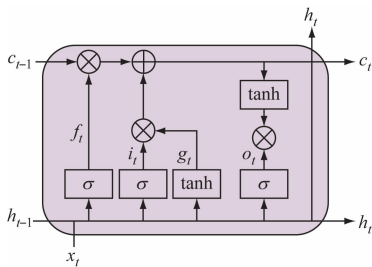


图 1 LSTM 基本结构

Fig. 1 Basic structure of LSTM

信息传递过程如下。

$$\begin{cases} f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \\ i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \\ g_t = \tanh(W_g \cdot [h_{t-1}, x_t] + b_g) \\ c_t = f_t * c_{t-1} + i_t * g_t \\ o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \\ h_t = o_t * \tanh(c_t) \end{cases} \quad (6)$$

式(6)中, σ 、 \tanh 均为激活函数; x_t 为第 t 个神经元的输入; W_f 、 b_f 分别为遗忘门的权重矩阵和偏置; W_i 、 b_i 分别为输入门 sigmoid 层的权重矩阵和偏置; W_g 、 b_g 分别为输入门 tanh 层的权重矩阵和偏置; W_o 、 b_o 分别为输出门的权重矩阵和偏置; f_t 为遗忘门的输出; i_t 、 g_t 分别为输入门中 sigmoid 层和 tanh 层的输出; c_{t-1} 、 c_t 分别为第 $t-1$ 个、第 t 个神经元的状态; h_{t-1} 、 h_t 分别为第 $t-1$ 个、第 t 个隐含层的输出; o_t 为输出门的输出。

2.5 组合模型概览

本文将本节所述模型组合形成二次分解深度学习模型, 模型结构见图 2。

为评估模型预测误差, 本文以均方根误差 (Root-Mean-Square Error, E_{RMS})、平均绝对百分比误差 (Mean Absolute Percentage Error, E_{MAPE})、均方根误差百分比 (Percentage Root Mean Square Error, E_{PRMS})、平均绝对误差百分比 (Percentage Mean Absolute Error, E_{PMA}) 和拟合优度 (Goodness of Fit, R^2) 为预测评价指标。

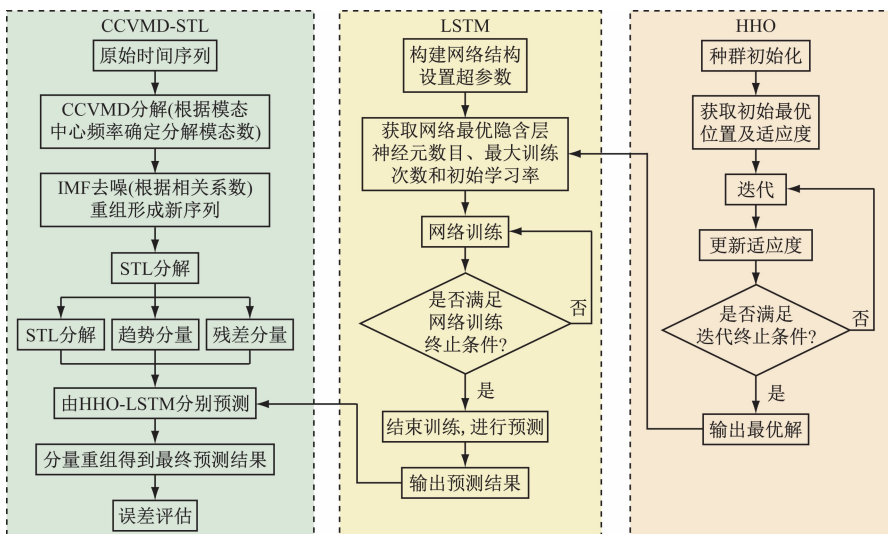


图 2 组合模型概览

Fig. 2 Overview of the combinatorial model

3 数据描述

本文选取世界第一大集装箱港——上海港自2014年1月至2024年2月的月度港口集装箱吞吐量数据共计122组,其中,2014年1月至2023年2月的110组数据为训练集(占样本总量90%),2023年3月至2024年2月的12组数据为测试集(占样本总量10%),数据来源为上海市统计局网站。港口集装箱吞吐量时间序列图见图3。

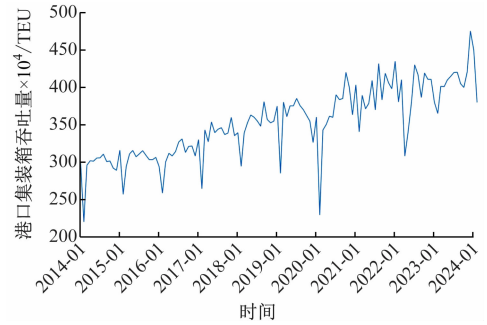


图3 港口集装箱吞吐量时间序列

Fig. 3 Port container throughput time series

4 结果和讨论

4.1 模型和参数

本文研究了7种单一模型 SARIMA、LSTM、BPNN、GM、RF、SVM、LSSVM 和6种组合模型 HHO-LSTM、SARIMA-HHO-LSTM、SARIMA-GM、VMD-HHO-LSTM、STL-HHO-LSTM 和 CCVMD-STL-HHO-

LSTM,在综合相关文献、建模经验和训练调试的基础上,设定各模型参数。

4.2 研究过程和结果

本文根据不同分解模态数下,各个分解模态的中心频率从低到高的情况,确定VMD的分解模态数 k ,见表1。

表1 分解模数及中心频率

Tab. 1 Decomposition modulus and central frequencies

k	中心频率								
2	3.1266×10^{-5}	0.1440							
3	2.9959×10^{-5}	0.0895	0.4178						
4	2.9825×10^{-5}	0.0884	0.4033	0.4845					
5	2.8648×10^{-5}	0.0795	0.1741	0.4119	0.4884				
6	2.8703×10^{-5}	0.0808	0.2381	0.3277	0.4179	0.4913			
7	2.8500×10^{-5}	0.0792	0.1723	0.3139	0.4162	0.4514	0.4946		
8	2.4494×10^{-5}	0.0335	0.0819	0.1835	0.3163	0.4162	0.4511	0.4946	

从表1可见,当 k 取值7时,分解模态出现最大中心频率0.4946,且相邻分解模态的中心频率变化率较大,模态区分度较好。当 k 取值8时,分解模态中心最大频率仍为0.4946,且第6模态和第7模态的中心频率较为接近,可能存在模态混叠。故确定分解模态数 k 为7。模态分解结果见图4。

经VMD处理的各个分解模态和原始时间序列之间的相关系数计算结果见表2。

为筛选包含原始时间序列重要特征信息的模态分量,同时过滤噪声和关联度较小的序列信息,达到较好的除噪效果,本文设定相关系数 R_i 的阈值为0.2,选取相关系数在阈值以上的分解模态序列 $F_{IM3}-F_{IM7}$ 加和重组,得到去噪后的新序列。

原始时间序列 S_T ,VMD分解序列 $F_{IM1}-F_{IM7}$,及去噪重组后的新序列 S_{T_clean} 详见表3。

将 S_{T_clean} 序列进行STL分解,由HHO-LSTM分别预测趋势项、季节项和不规则项,将预测结果重

组,得到最终的模型预测序列。HHO算法对分解序列预测的优化过程及相关预测结果见表4~表5、图5~图7。

本文按照表2给定参数,用其余12种模型对上海港集装箱吞吐量进行预测。所有模型对测试集数据的预测曲线见图8,模型预测性能见表6。

从表6可见,13种预测模型中,3种分解组合模型的整体预测效果显著优于其他模型,见图9。

4.3 结果讨论

1) 关于CCVMD-STL-HHO-LSTM模型的讨论

从图3直观地可见,上海港集装箱吞吐量序列整体上存在显著的非平稳性和波动性,且每年2月出现全年集装箱吞吐量月度最低值,集装箱吞吐量存在明显季节性特征。

从图6看,HHO-LSTM对经STL处理得到的季节项和趋势项的预测效果较为理想,整体上拟合较好,说明经HHO算法搜索能够获取有效的网络初

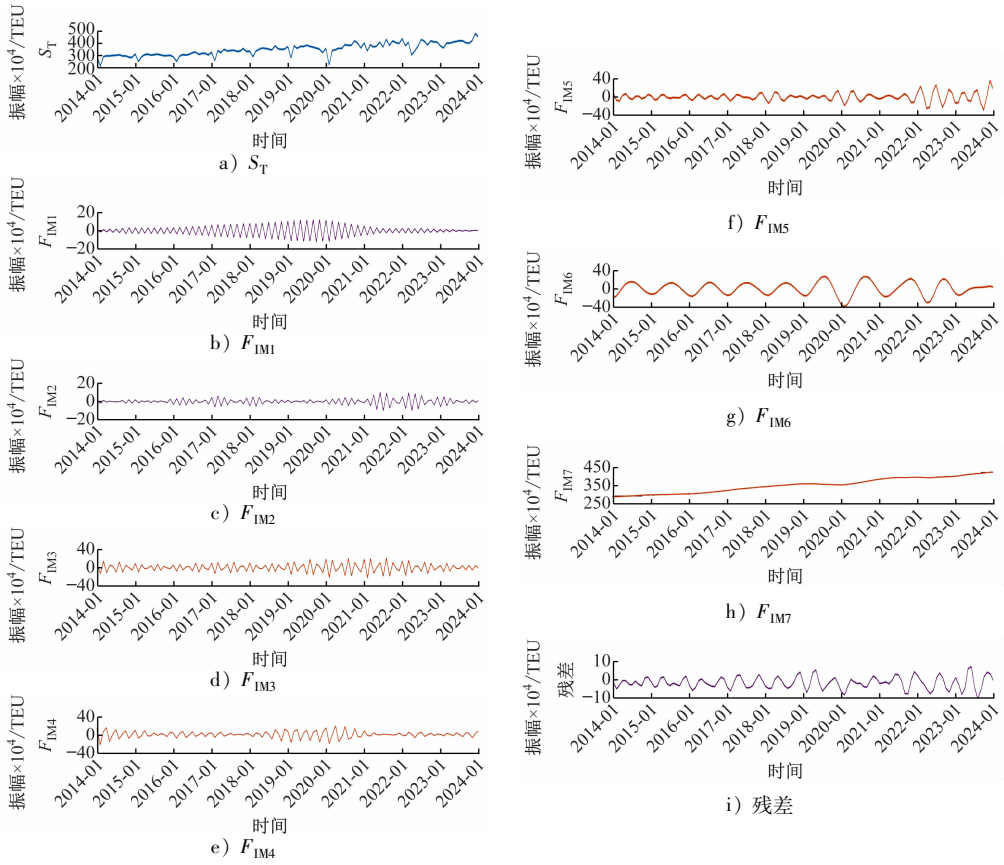


图 4 VMD 分解结果

Fig. 4 Results of VMD

表 2 各分解模态相关系数

Tab. 2 Correlation coefficient for IMFs

分解模态	F_{IM1}	F_{IM2}	F_{IM3}	F_{IM4}	F_{IM5}	F_{IM6}	F_{IM7}
相关系数 R_i	0.158 9	0.140 3	0.242 5	0.221 9	0.267 6	0.345 7	0.846 9

表 3 CCVMD 分解序列

Tab. 3 CCVMD decomposition sequences

序号	S_T	F_{IM1}	F_{IM2}	F_{IM3}	F_{IM4}	F_{IM5}	F_{IM6}	F_{IM7}	S_{T_clean}
1	299.72	0.604 735 334	0.903 305 365	5.307 437 702	14.435 051 58	7.746 028 404	-19.973 089 72	289.922 136 4	297.437 564 4
2	220.59	-1.427 798 019	-1.966 299 603	-13.203 818 93	-25.426 123 53	-7.116 455 137	-16.257 358 43	289.769 949 2	227.766 193 1
3	296.26	1.321 526 437	1.389 627 903	14.745 725 35	6.067 450 99	-8.323 873 797	-7.898 667 711	290.541 59	295.132 224 9
4	302.25	-1.362 818 445	-0.732 714 672	-11.444 461 35	16.380 461 01	4.292 670 079	2.088 969 923	291.675 953 7	302.993 593 4
5	301.68	1.756 949 4	0.511 821 266	5.170 226 034	-16.800 520 13	7.347 512 893	9.929 615 179	292.478 166 2	298.125 000 1
6	305.89	-1.960 483 993	-0.124 289 014	2.778 941 386	0.800 458 131	-1.972 348 396	14.194 451 52	293.096 11	308.897 612 6
7	306.17	1.916 383 294	-0.370 734 381	-9.785 417 541	11.418 709 1	-5.066 450 628	15.310 376 95	294.052 810 4	305.930 028 2
8	311.09	-2.278 110 146	-0.009 988 782	12.392 842 3	-10.662 623 78	1.556 839 322	13.571 709 75	295.431 424 8	312.290 192 4
9	301.13	2.964 082 566	1.100 411 921	-10.097 270 86	0.839 712 363	2.139 774 846	8.465 360 775	296.527 368 7	297.874 945 8
10	301.98	-3.409 909 802	-1.906 050 404	5.101 037 435	8.651 634 478	-3.388 818 416	1.485 503 455	297.388 116	309.237 473
11	292.34	3.408 673 061	1.970 437 747	0.418 776 049	-9.409 288 718	-0.911 000 648	-4.658 822 588	298.563 321 7	284.002 985 8
12	289.46	-3.270 622 378	-1.726 186 483	-5.398 926 76	0.164 828 316	5.811 506 1	-9.204 763 106	299.642 204 8	291.014 849 3
...
111	401.63	1.469 202 873	-4.676 884 92	9.861 451 13	2.797 489 816	7.692 136 633	-12.576 171 45	404.273 273 6	412.048 179 7
112	401.2	-1.208 634 657	3.790 821 54	-8.089 193 13	0.187 905 967	10.338 866 01	-7.556 673 805	406.928 444 2	401.809 349 2
113	410.17	1.223 530 064	-2.330 481 41	5.348 563 376	-2.718 691 891	-6.509 410 442	-2.596 286 661	409.057 080 6	402.581 255
114	415.01	-0.827 593 377	1.349 800 392	-1.453 670 977	4.029 133 942	-9.171 991 87	0.632 507 626	410.700 326 8	404.736 305 5
115	420.13	0.584 861 049	-0.254 781 85	-2.086 147 816	-0.485 836 424	10.435 374 02	2.261 309 509	412.248 195 5	422.372 894 8
116	420.45	-0.557 769 41	-0.950 682 614	4.153 769 933	-4.971 963 076	15.679 243 17	2.974 049 715	413.934 219 8	431.769 319 6
117	405.16	0.520 164 562	1.820 052 477	-4.823 930 265	4.828 781 103	-11.318 785 01	3.034 852 141	415.525 463 6	407.246 381 6
118	400.12	-0.471 977 347	-2.204 818 549	4.894 387 034	1.866 743 189	-28.046 915 87	3.288 334 14	417.108 491 8	399.111 040 3
119	420.74	0.055 512 205	1.650 813 474	-5.254 623 951	-7.722 750 437	1.894 583 55	4.593 107 453	419.068 437 7	412.578 754 3
120	475.53	0.363 120 423	-0.864 540 212	4.739 450 035	3.088 673 172	36.661 506 01	5.453 922 312	420.824 765	470.768 316 5
121	450.17	0.266 453 002	1.295 929 671	-1.885 009 533	8.307 992 25	16.481 841 96	3.695 195 734	421.255 112 6	447.855 133
122	380.17	-0.283 166 717	-0.734 964 174	0.199 100 414	-6.380 735 804	-31.567 515	0.997 255 109	420.797 258 2	384.045 362 9

注:因篇幅限制,表中仅呈现部分数据。

表 4 STL 分解序列及模型预测结果
Tab.4 STL decomposition sequences and model predictions

序号	S_{T_clean}	不规则项	季节项	趋势项	不规则项-预测	季节项-预测	趋势项-预测	模型预测序列
1	297.437 564 4	1.053 25	1.003 37	281.450 23	-	-	-	-
2	227.766 193 1	0.916 48	0.874 03	284.339 83	-	-	-	-
3	295.132 224 9	1.020 23	0.997 1	290.119 03	-	-	-	-
4	302.993 593 4	1.022 64	1.001 89	295.727 02	1.024 776	1.002 147 1	296.286 29	304.278 996 8
5	298.125 000 1	1.001 38	0.991 65	300.221 7	0.998 503 57	0.991 923 45	301.058 81	298.180 418 6
6	308.897 612 6	1.003 01	1.028 18	299.531 85	0.970 975 28	1.017 402 5	304.583 34	300.889 559 5
7	305.930 028 2	1.000 37	1.025 28	298.275 82	0.989 712 3	1.015 186 5	302.807 98	304.244 066 8
8	312.290 192 4	0.989 39	1.062 14	297.174 01	0.997 923 55	1.060 914 9	300.727 39	318.383 685 5
9	297.874 945 8	1.012 58	0.993 64	296.056 8	1.008 295 2	0.997 215 87	299.731 29	301.376 207 9
10	309.237 473	1.004 49	1.045 64	294.416 62	1.003 840 3	1.045 471 9	298.821 08	313.608 786 7
11	284.002 985 8	0.975 12	0.988 65	294.594 03	0.989 919 42	0.990 213 22	297.352 66	291.474 383 4
12	291.014 849 3	0.990 3	0.988 43	297.305 37	1.012 181 3	0.992 226 78	297.776 61	299.061 037 3
...
111	412.048 179 7	1.016 51	0.997 1	406.531 03	1.018 058 2	0.996 313 15	404.244 72	410.027 348 6
112	401.809 349 2	0.988 73	1.001 89	405.623 29	0.988 167 94	1.002 147 1	406.761 02	402.811 222 2
113	402.581 255	1.005 85	0.991 65	403.612 75	0.984 359 5	0.991 923 45	405.075 38	395.519 356 6
114	404.736 305 5	0.977 5	1.028 18	402.706 46	1.002 951 6	1.017 402 5	402.168 12	410.374 547 6
115	422.372 894 8	1.015 23	1.025 28	405.778 66	1.017 442 9	1.015 186 5	401.400 21	414.603 987 6
116	431.769 319 6	1.005 52	1.062 14	404.275 78	1.012 181 3	1.060 914 9	406.039 28	436.020 498 8
117	407.246 381 6	1.014 7	0.993 64	403.914 23	0.990 984 74	0.997 215 87	404.073 15	399.315 475 4
118	399.111 040 3	0.932 92	1.045 64	409.132 82	0.980 737 09	1.045 471 9	402.885 01	413.091 323 7
119	412.578 754 3	0.982 16	0.988 65	424.899 15	1.017 918 5	0.990 213 22	410.143 25	413.406 495 5
120	470.768 316 5	1.077 76	0.988 43	441.917 58	1.055 983 3	0.992 226 78	430.212 49	450.765 852 8
121	447.855 133	0.983 13	1.003 37	454.008 13	0.985 964 66	1.005 813 5	447.383 27	443.668 452 3
122	384.045 362 9	0.955 09	0.874 03	460.053 41	0.947 873 12	0.886 019 17	456.132 72	383.075 655

注:因篇幅限制,表中仅呈现部分数据。

表 5 HHO 算法返回最优参数

Tab.5 Optimal parameters returned by HHO algorithm

参数	不规则项序列	季节项序列	趋势项序列
隐含层神经元数量/个	63	98	42
最大训练次数/次	801	1 965	771
初始学习率	0.100 0	0.097 8	0.056 1

始超参数,提升了 LSTM 对输入输出数据间映射关系的构建能力,改善了预测准确性。由于不规则项数据内部不存在紧密关联,故 HHO-LSTM 对不规则项的预测效果稍显不足。

从图 7 看,相较于原始序列,经 CCVMD 去噪处理后的序列的局部极值有所削弱,除对极值点的拟

合略显不足外,由 CCVMD-STL-HHO-LSTM 模型最终得到的预测曲线整体上能较好反映原始序列的特征信息。

2) 模型预测结果比较

本文研究了 13 种预测模型,从图 8 展示的结果看,SARIMA、SARIMA-GM 和 SARIMA-HHO-LSTM 三种模型的预测曲线和原始序列曲线存在显著偏离,预测结果不理想。LSTM、BPNN、GM、RF、SVM、LSSVM 和 HHO-LSTM 等模型的预测曲线和原始序列曲线的偏离相较于 SARIMA 相关模型小,但也没能反映原始序列“涨-落-涨-落”的走势。

VMD-HHO-LSTM、STL-HHO-LSTM 和本文提出的 CCVMD-STL-HHO-LSTM 三种模型的预测曲线在整体拟合方面优于前述模型。其中,STL-HHO-LSTM 模型对原始序列整体走势的拟合表现较好,

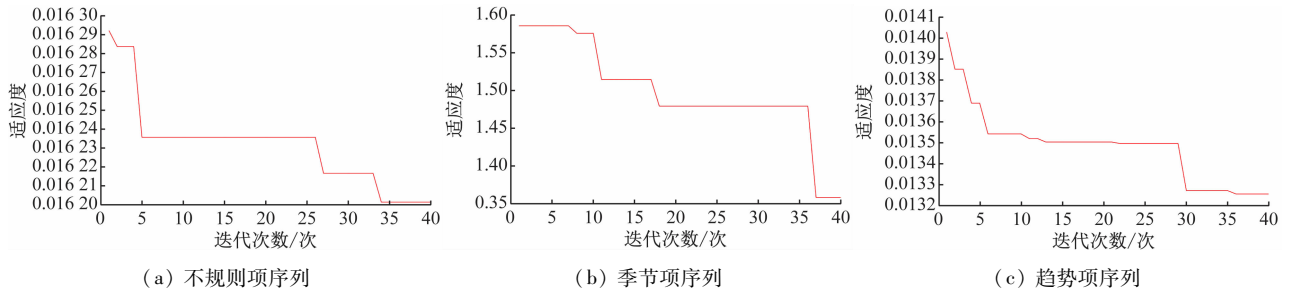


图5 HHO算法迭代曲线

Fig.5 Iterative curves for HHO algorithm

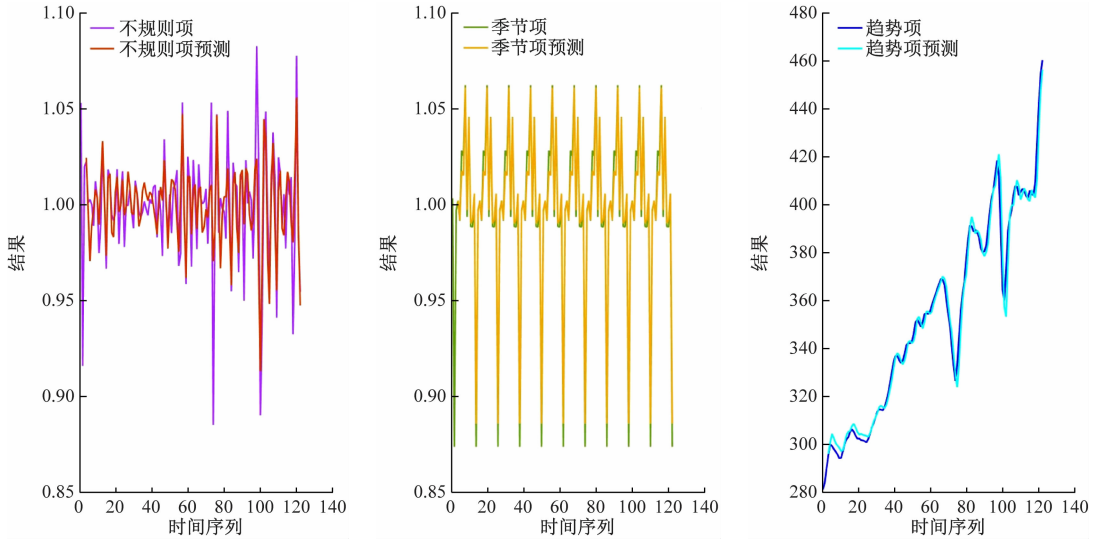


图6 STL分解与预测

Fig.6 STL decomposition and forecasting

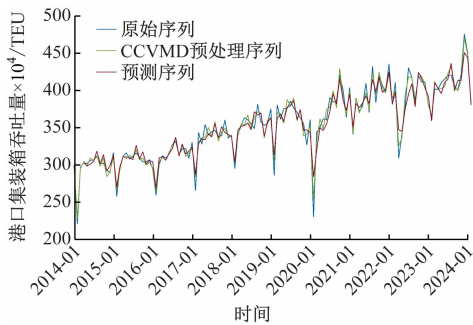


图7 CCVMD-STL-HHO-LSTM模型的相关序列

Fig.7 Sequences about CCVMD-STL-HHO-LSTM

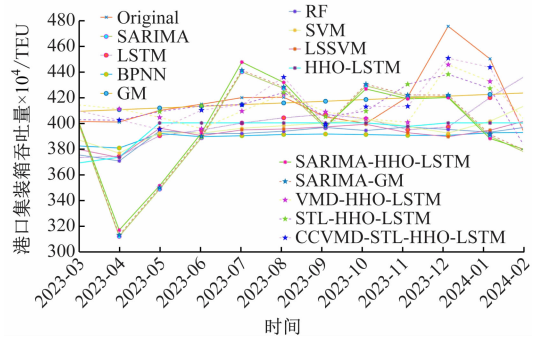


图8 各种模型预测曲线

Fig.8 Forecast curves of models

但在捕捉跟踪序列极值的能力上略显不足;VMD-HHO-LSTM在走势跟踪上逊色于STL-HHO-LSTM,但在极值预测(如2023年12月的PCT)上更接近原始序列;这两种模型的预测效果均明显优于HHO-LSTM,可见数据预处理带来的降噪作用能显著提高模型预测性能。CCVMD-STL-HHO-LSTM继承了CCVMD分解处理的降噪优势和STL分解的捕捉序列特征优势,特别是在原始序列存在较大跃跌时,拟合极值的能力较好。

从表6评价指标结果看,7种单一模型均存在较大预测误差,预测效果较差,3种简单组合模型也没有显著改善预测结果。特别地,表6中前10种模型的拟合优度 R^2 均为负数或十分接近0,说明模型对时间序列自身信息的识别、利用不足,模型预测的可解释性差,对比后3种分解组合模型的拟合优度可知,对时间序列数据进行分解预处理可大幅提升模型的信息捕捉能力,增强可解释性。3种分解组合模型的评价指标数据均显著优于其他模型,其中,

表 6 模型评价指标值

Tab. 6 Values of evaluation indicators for models

模型名称	评价指标				
	E_{RMS}	E_{MAP}	E_{PRMS}	E_{PMA}	R^2
SARIMA	41.124 2	0.071 2	0.098 7	0.072 2	-1.950 3
LSTM	34.402 8	0.065 074	0.082 559	0.065 877	-1.064 7
BPNN	35.595 4	0.067 174	0.085 421	0.069 505	-1.210 4
GM	22.945 5	0.037 002	0.055 064	0.037 566	0.081 525
RF	34.670 5	0.066 068	0.083 201	0.068 104	-1.097
SVM	33.811 8	0.062 607	0.081 141	0.064 388	-0.994 4
LSSVM	35.394 5	0.065 353	0.084 939	0.067 5	-1.185 5
HHO-LSTM	31.528 2	0.057 468	0.075 66	0.059 254	-0.734 09
SARIMA-HHO-LSTM	40.305	0.0709 8	0.096 72	0.072 16	-1.833 9
SARIMA-GM	40.522 3	0.070 67	0.097 244	0.071 65	-1.864 6
VMD-HHO-LSTM	14.894 9	0.029 816	0.035 744	0.030 295	0.612 97
STL-HHO-LSTM	13.549 6	0.019 153	0.032 516	0.020 203	0.679 72
CCVMD-STL-HHO-LSTM	11.189 2	0.021 703	0.026 852	0.022 14	0.781 59

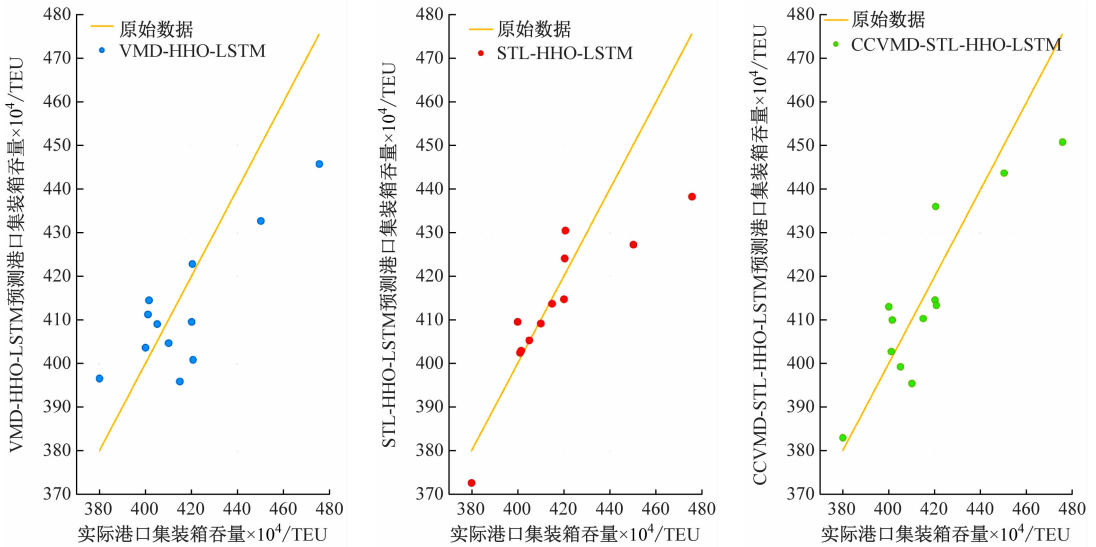


图 9 分解组合模型预测结果

Fig. 9 Predictions of decomposed combinatorial models

CCVMD-STL-HHO-LSTM 在所有指标上均优于 VMD-STL-HHO-LSTM,在 E_{RMS} 、 E_{PRMS} 和 R^2 上均优于 STL-HHO-LSTM,在 E_{MAP} 和 E_{PMA} 上略高于 STL-HHO-LSTM,并且在所有模型中具备相对最高的拟合优度和相对最低的预测误差,可见本文提出的二次分解模型融合了 CCVMD 和 STL 对数据预处理的优势,在预测综合性能上比未经数据预处理的简单组合模型和一次分解组合模型有更好表现。

3) 与其他研究预测结果的比较

由于不同研究的测试集数据量不同,本文以相对指标 E_{MAP} 为横向比较指标,就所提出的 CCVMD-STL-HHO-LSTM 模型和已有研究模型的预测效果进行比较分析,见表 7。

从上述分析可见,本文提出的 CCVMD-STL-HHO-LSTM 模型在 MAPE 指标上优于部分研究的推荐模型,存在一定的预测优势。

表7 本文推荐模型与其他研究模型预测结果比较

Tab.7 Comparison of the prediction results of the recommended model in this paper with others

序号	类别	模型名称	E_{MAP}	备注
1		CCVMD-STL-HHO-LSTM	0.021 7	本文
2	上海港集装箱 吞吐量预测的 有关研究	VMD-SARIMA-卷积神经网络-双向长短期 记忆网络-注意力机制	0.041 9	参考文献[17]
3		经验小波变换-时域卷积网络-核最小平方 误差	0.022 2	参考文献[19]
4		LSTM	0.032 7	参考文献[23]
5		灰色关联分析-Choquet 模糊积分	0.028 5	参考文献[24]
6	其他港口(集装箱) 吞吐量预测的 有关研究	RF	0.926 7	参考文献[5]
7		SVR	0.023 8	参考文献[9]
8		RF	0.043 4	参考文献[25]
9		LSTM-船舶自动识别系统	0.107 0	参考文献[26]

5 结束语

在集装箱运输在海运占比越来越大的趋势下,准确预测港口集装箱吞吐量对港口航运企业和政府管理部门研判贸易形势、制定发展规划、调整发展政策具有重要意义。基于港口集装箱吞吐量非线性、非平稳的特点,本文提出了基于二次分解的新型深度学习模型 CCVMD-STL-HHO-LSTM,以结合相关系数的 VMD 方法对原始序列除噪,经 STL 处理实现二次分解,再通过 HHO 算法优化 LSTM 超参数,从而提升模型的预测性能。以上海港集装箱吞吐量为学习与预测对象,比较该模型与其余 12 种模型和既有研究提出的部分模型的预测效果。研究结果表明。

1) 面对波动性较大的数据序列时,在预测性能上,机器学习方法整体优于传统模型,组合模型方法整体优于对应的单一模型,对数据进行预处理的分解组合模型显著优于不做数据预处理的简单组合模型。

2) 结合相关系数的变分模态分解方法降低了分解模态筛选时的主观经验干扰,具有更好的可靠性。CCVMD 和 STL 相结合的二次分解方式,继承了 VMD 的除噪优势和 STL 的数据特征捕捉优势,有助于减少人为主观经验干扰,筛选出序列重要特征信息。

3) HHO 算法寻优能有效提升 LSTM 的学习性能。相较于比较模型和部分研究推荐模型,本文提出的 CCVMD-STL-HHO-LSTM 在 5 种评价指标上的整体表现最优。

值得注意的是,虽然本文借助相关系数筛选分解模态,但在 VMD 分解阶段对分解模态数等参数的确定方法,业内尚无统一定论。因此,在今后的研究工作中,通过理论推导和大量实证分析,探讨客观科学的分解参数选取方法,将有助于进一步提高预测的准确性。

参 考 文 献

- [1] The World Bank. The container port performance index 2023: a comparable assessment of performance based on vessel time in port (Fine) [R]. Washington, DC: World Bank, 2024.
- [2] MUNIM Z H. Does higher technical efficiency induce a higher service level? A paradox association in the context of port operations [J]. The Asian Journal of Shipping and Logistics, 2020, 36(4): 157-168.
- [3] ESKAFI M, KOWSARI M, DASTGHEIB A, et al. A model for port throughput forecasting using Bayesian estimation [J]. Maritime Economics & Logistics, 2021, 23: 348-368.
- [4] GOSASANG V, YIP T L, CHANDRAPRAKAIKUL W. Long-term container throughput forecast and equipment planning: the case of Bangkok Port [J]. Maritime Business Review, 2018, 3(1): 53-69.
- [5] AWAH P C, NAM H, KIM S. Short term forecast of container throughput: new variables application for the port of douala [J]. Journal of Marine Science and Engineering, 2021, 9, 720.
- [6] WANG J Z, SHAO Y Y, JIANG H, et al. A multi-variable hybrid system for port container throughput deterministic and uncertain forecasting [J]. Expert

- Systems with Applications, 2024, 237: 121546.
- [7] 吴丽淑. 基于 CS-VMD-LSTM 的珠三角典型港口货物吞吐量预测研究[D]. 大连:大连海事大学,2023.
WU L S. Research on typical ports cargo throughput prediction in pearl river delta based on CS-VMD-LSTM [D]. Dalian: Dalian Maritime University, 2023. (in Chinese)
- [8] HUANG D, GRIFOLL M, SANCHEZ-ESPIGARES J A, et al. Hybrid approaches for container traffic forecasting in the context of anomalous events: the case of the Yangtze River Delta region in the COVID-19 pandemic [J]. Transport Policy, 2022, 128: 1-12.
- [9] CHAN H K, XU S J, QI X G. A comparison of time series methods for forecasting container throughput[J]. International Journal of Logistics Research and Applications, 2019, 22 (3): 294-303.
- [10] SETIAWAN R, SARNO R, FAUZAN A C, et al. Evaluation of container forecasting methods for analyzing port container terminal performance using agent-based simulation[C]. International Conference on Information and Communications Technology (ICOIACT), 2018: 286-291.
- [11] KOURENTZES N, BARROW D, PETROPOULOS F. Another look at forecast selection and combination: evidence from forecast pooling[J]. International Journal of Production Economics, 2019, 209: 226-235.
- [12] WANG X Q, HYNDMAN R J, LI F, et al. Forecast combinations: an over 50-year review[J]. International Journal of Forecasting, 2023, 39(4): 1518-1547.
- [13] ZHANG Y F, FU Y H, LI G H. Research on container throughput forecast based on ARIMA-BP neural network [J]. Journal of Physics: Conference Series, 2020, 1634: 012024.
- [14] 孙晓聪, 付玉慧. 基于 RF-双向 LSTM 的集装箱吞吐量预测 [J]. 上海海事大学学报, 2022, 43 (1): 60-65.
SUN X C, FU Y H. Container throughput prediction based on RF-bidirectional LSTM [J]. Journal of Shanghai Maritime University, 2022, 43(1): 60-65.
- [15] DRAGAN D, KESHAVARZSALEH A, INTIHAR M, et al. Throughput forecasting of different types of cargo in the adriatic seaport Koper [J]. Maritime Policy & Management, 2021, 48: 1, 19-45.
- [16] DU P, WANG J Z, YANG W D, et al. Container throughput forecasting using a novel hybrid learning method with error correction strategy [J]. Knowledge-Based Systems, 2019, 182: 104853.
- [17] XIAO Y, XUE X F, HU Y, et al. Novel decomposition and ensemble model with attention mechanism for container throughput forecasting at four ports in Asia [J]. Transportation Research Record, 2023, 2677(6): 530-547.
- [18] 肖进, 文章, 刘博, 等. 基于选择性深度集成的集装箱吞吐量混合预测模型研究[J]. 系统工程理论与实践, 2022, 42(4): 1107-1128.
XIAO J, WEN Z, LIU B, et al. A hybrid model based on selective deep-ensemble for container throughput forecasting[J]. Systems Engineering-Theory & Practice, 2022, 42(4): 1107-1128.
- [19] XIAO Y, XIE M H, HU Y, et al. Effective multi-step ahead container throughput forecasting under the complex context [J]. Journal of Forecasting, 2023, 42: 1823-1843.
- [20] SANGURI K, SHANKAR S, PUNIA S, et al. Hierarchical container throughput forecasting: the value of coherent forecasts in the management of ports operations [J]. Computers & Industrial Engineering, 2022, 173: 108651.
- [21] ZHANG D, LI X Y, WAN C P, et al. A novel hybrid deep-learning framework for medium-term container throughput forecasting: an application to China's Guangzhou, Qingdao and Shanghai hub ports [J]. Maritime Economics & Logistics, 2024, 26: 44-73.
- [22] SHEN Y F, ZHENG W, YIN W J, et al. Feature extraction algorithm using a correlation coefficient combined with the VMD and its application to the GPS and GRACE [C]. IEEE Access, 2021, 9: 17507-17519.
- [23] 王凤武, 张晓博, 阎际驰, 等. 基于 LSTM 的上海港集装箱吞吐量预测 [J]. 中国航海, 2022, 45 (2): 109-114.
WANG F W, ZHANG X B, YAN J C, et al. Prediction of container throughput of Shanghai Port with LSTM [J]. Navigation of China, 2022, 45 (2): 109-114. (in Chinese)
- [24] WU G, HU Y C, CHIU Y J, et al. Forecast combination using grey relational analysis and Choquet fuzzy integral for container throughput forecasting [J]. Expert Systems with Applications, 2024, 252: 124170.
- [25] 谢新连, 王余宽, 许小卫, 等. 基于随机森林算法的港口集装箱吞吐量预测方法 [J]. 重庆交通大学学报 (自然科学版), 2022, 41(2): 15-20.
XIE X L, WANG Y K, XU X W, et al. Port container throughput forecasting method based on random forest algorithm [J]. Journal of Chongqing Jiaotong University (Natural Science), 2022, 41(2): 15-20. (in Chinese)
- [26] NAKASHIMA M, SHIBASAKI R. Can AIS data improve the short-term forecast of weekly dry bulk cargo port throughput? -a machine-learning approach [J]. Maritime Policy & Management, 2023.