

DOI:10.19322/j.cnki.issn.1006-4710.2023.03.004

<https://xuebao.xaut.edu.cn>

引文格式:赵峰,徐丹华. 基于改进 HHO 算法的碳交易价格组合预测研究[J]. 西安理工大学学报,2023,39(3):330-338.

ZHAO Feng, XU Danhua. Research on the prediction of carbon trading price combination by the improved HHO algorithm[J]. Journal of Xi'an University of Technology, 2023, 39(3): 330-338.

基于改进 HHO 算法的碳交易价格组合预测研究

赵 峰, 徐丹华

(安徽工业大学 管理科学与工程学院, 安徽 马鞍山 243032)

摘要: 为提高碳交易价格预测精度,建立了多策略改进哈里斯鹰算法的碳交易价格组合预测模型。一方面对碳交易价格序列的高频和低频序列分别建立 ARIMA 模型和指数平滑模型,通过加和对碳交易价格进行预测。另一方面综合考虑碳交易价格的经济指标和技术指标,通过 Pearson 相关系数筛选出 6 个与下一日碳交易价格高度相关的变量作为解释变量,建立多策略改进哈里斯鹰优化极限学习机模型(THHO_ELM)。最后,对模型 I 和模型 II 建立基于 l_p 范数的组合预测模型。结果表明,组合预测模型优于单一的分类模型。

关键词: 碳价格预测; 完全自适应噪声集合经验模态分解; 哈里斯鹰优化算法; 极限学习机; l_p 范数
中图分类号: F832.5; X196 **文献标志码:** A **文章编号:** 1006-4710(2023)03-0330-09

Research on the prediction of carbon trading price combination by the improved HHO algorithm

ZHAO Feng, XU Danhua

(School of Management Science and Engineering, Anhui University of Technology, Maanshan 243032, China)

Abstract: In order to improve the prediction accuracy of carbon trading price, a multi strategy improved Harris hawk algorithm for carbon trading price combination prediction is established. On the one hand, ARIMA model and exponential smoothing model are established for the high-frequency and low-frequency series of carbon trading price series respectively, and the carbon trading price is predicted by adding. On the other hand, considering the economic and technical indicators of carbon trading price, six variables highly related to the next day's carbon trading price are selected as explanatory variables through the Pearson correlation coefficient, and a multi strategy improved Harris hawk optimization extreme learning machine model is established (THHO_ELM). Finally, Model I and Model II are combined based on the l_p norm. The results show that the combined prediction model is better than the single classification model.

Key words: carbon price forecast; CEEMDAN decomposition; Harris hawk optimization algorithm; extreme learning machine; l_p norm

我国碳交易已经拥有一定规模且市场前景广阔^[1-2]。碳价格在促进“双碳”目标实现的基础上具有一定的负面影响^[3],如何通过碳交易价格引导投资者更好地利用碳排放权交易市场进行投资,推动市场的理性发展,已成为问题的关键。

目前,国内外学者不断对碳交易价格及其影响因素进行研究和分析,并尝试建立模型对碳交易价格进行预测。有关碳交易价格预测的研究主要涉及两个方面。

1) 针对碳交易价格的时间序列建立预测模型。

收稿日期: 2022-09-05; 网络首发日期: 2022-12-13

网络首发地址: <https://kns.cnki.net/kcms/detail//61.1294.N.20221212.1737.002.html>

基金项目: 国家自然科学基金资助项目(71872002);安徽省高校人文社会科学研究重点项目(SK2019A0072)

第一作者: 赵峰,男,硕士,副教授,研究方向为数据挖掘、战略管理。E-mail: 345499343@qq.com

通信作者: 徐丹华,女,硕士生,研究方向为数据挖掘。E-mail: 2325821724@qq.com

文献[4]~[8]通过对碳价时间序列进行分解、重组,分别对重组后的序列建立不同的模型进行碳交易价格预测。Zhang 等^[9]提出了一种先进的深度神经网络模型来预测碳价格。Xu 等^[10]利用时间序列复杂网络分析技术和极限学习机算法提出了一种新的碳价预测模型。Yang 等^[11]提出了包括改进数据特征提取技术,建立 LSTM、CNN、ELM 预测子模型及多目标优化算法加权策略的集成预测系统(EPS)。彭武元等^[12]针对碳价格时间序列,建立了马尔科夫转换多重分形模型。Zhang 等^[13]提出了一种整合 PSO 和多输出支持向量回归(MSVR)的混合方法来预测碳价格。蒋锋等^[14]基于混沌粒子群优化(CPSO)算法提出了一种优化的 BP 神经网络碳价预测模型。

2) 针对碳价波动的相关因素建立预测模型。Xie 等^[15]创新性地使用文本在线新闻来构建与气候相关的变量,并结合其它变量利用长短期记忆网络和随机森林模型对碳价进行预测。刘金培等^[16]对非结构化数据、结构化数据和碳交易价格分别进行分解、重组,建立了自回归积分滑动平均模型、偏最小二乘回归和神经网络模型。魏宇等^[17]对影响我国碳价格的相关因素进行了分析,筛选出 9 个重要影响因素,并对各类经典预测模型进行对比分析。

通过相关文献分析发现,针对国内碳交易价格预测的文献相对较少,且大部分文献仅从碳交易价格本身或碳交易价格相关影响因素着手进行建模预测,目前尚没有综合考虑两种建模方法的相关文献。因此,本文综合考虑以上两个方面,分别建立 CEEMDAN-ARIMA-指数平滑模型和改进的哈里斯鹰优化极限学习机的碳价预测模型,并通过 l_p 范数建立组合预测模型。为此,论文在现有文献的基础上,构建了基于 l_p 范数的碳交易价格组合预测模型。该模型具有如下创新性:

1) 为了充分挖掘碳交易价格所带来的信息,一方面考虑碳交易价格本身建立了 CEEMDAN-ARIMA-指数平滑模型,另一方面考虑碳交易价格的经济指标和技术指标等建立了 THHO-ELM 模型;

2) 利用多策略融合改进的哈里斯鹰算法优化极限学习机模型中的权值和阈值,并利用优化后的模型对广东省碳交易价格进行训练和预测;

3) 为充分利用上述两个模型的优势,建立了基于 l_p 范数的误差平方绝对值模型。实证结果表明,本文提出的 l_p 范数组合预测模型优于单一的分类模型,对碳交易价格预测精度较高。此外,本文在我国的东、中、西部分别选取上海、湖北和重庆碳交易试点进行预测,进一步验证了模型的可靠性。

1 理论基础

1.1 完全自适应噪声集合经验模态分解

经验模态分解(empirical mode decomposition, EMD)是处理非平稳信号的方法之一,其本质是将原始的复杂信号分解成包含原始信号部分特性的多个 IMF 分量和残余分量之和^[18]。EMD 算法虽对信号具有一定的适应性,但易产生模态混叠现象,从而影响信号分解质量。而完全自适应噪声集合经验模态分解(CEEMDAN)加入了经 EMD 分解后含辅助噪声的 IMF 分量,并对分解得到的 IMF 分量进行总体平均,完美解决了 EMD 算法中存在的模态混叠现象。

CEEMDAN 算法的实现步骤:

Step1:在原始分解信号 $y(t)$ 中加入符合标准正态分布的高斯白噪声信号 $v^j(t)$,得到新信号 $y(t) + v^j(t)$ 。对其进行 EMD 分解,得到第 1 个本征模态分量 C_1 。对产生的 N 个模态分量进行平均,得到 CEEMDAN 分解的第 1 个本征模态分量 $\overline{C_1(t)}$,其中 $\overline{C_1(t)} = \frac{1}{N} \sum_{j=1}^N C_1^j(t)$ 。

Step2:从原始分解信号 $y(t)$ 中去除第一个本征模态分量 $\overline{C_1(t)}$ 可得到第一个模态分量后的残差 $r_1(t)$,将残差 $r_1(t)$ 作为原始分解信号重复 Step1,即可得到 CEEMDAN 分解的第 2 个本征模态分量 $\overline{C_2(t)}$ 。

Step3:将新的残差作为原始分解信号重复 Step1,当其为单调函数,不能继续进行分解时,算法结束。最终原始信号 $y(t)$ 被分解为: $y(t) = \sum_{k=1}^K \overline{C_k(t)} + r_k(t)$, K 为本征模态分量个数。

1.2 哈里斯鹰优化算法

1.2.1 传统的哈里斯鹰优化算法

哈里斯鹰算法(harris hawks optimization, HHO)是基于哈里斯鹰的群体捕食行为而提出的一种元启发式算法^[19],主要包括探索阶段、探索到开发的转换以及局部开采三个阶段。每个阶段的具体描述如下。

阶段 1:探索阶段哈里斯鹰的位置更新:

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

式中: q, r_1, r_2, r_3, r_4 均为 $[0, 1]$ 内的随机数; ub, lb 为搜索空间的上、下界; X_{rand} 为群体内随机个体位

置; X_{rabbit} 为猎物位置; X_m 为种群内所有个体的平均位置。

阶段 2: 探索到开发的转换:

$$E = 2E_0(1 - \frac{t}{T}) \quad (2)$$

式中: E_0 为初始状态的逃逸能量, 在算法迭代过程中于 $[-1, 1]$ 之间随机变化; t 为种群当前迭代次数; T 为种群的最大迭代次数。

阶段 3: HHO 算法根据哈里斯鹰的围捕方式和猎物的逃脱行为, 提出了四种攻击策略。

策略 1: 当逃逸能量 $|E| \geq 0.5$ 且随机数 $r \geq 0.5$ 时, 哈里斯鹰采取软围攻策略。其位置更新为:

$$X(t+1) = X_{\text{rabbit}}(t) - E|JX_{\text{rabbit}}(t) - X(t)| \quad (3)$$

式中: $J = 2(1 - r_5)$ 为猎物的跳跃距离; r_5 为 $[0, 1]$ 内的随机数。

策略 2: 当逃逸能量 $|E| < 0.5$ 且随机数 $r \geq 0.5$ 时, 哈里斯鹰采取硬围攻策略。其位置更新为:

$$X(t+1) = X_{\text{rabbit}}(t) - E|X_{\text{rabbit}}(t) - X(t)| \quad (4)$$

策略 3: 当逃逸能量 $|E| \geq 0.5$ 且随机数 $r < 0.5$ 时, 哈里斯鹰采取渐进式快速俯冲软包围。其位置更新为:

$$X(t+1) = \begin{cases} Y: X_{\text{rabbit}}(t) - E|JX_{\text{rabbit}}(t) - X(t)|, \\ \quad \text{if } F(Y) < F(X(t)) \\ Z: Y + S \times \text{LF}(D), \text{ if } F(Z) < F(X(t)) \end{cases} \quad (5)$$

式中: S 为随机行向量; D 为解决问题的维度; LF 为 levy 函数。

策略 4: 当逃逸能量 $|E| < 0.5$ 且随机数 $r < 0.5$ 时, 采取渐进式快速俯冲硬包围。其位置更新为:

$$X(t+1) = \begin{cases} Y: X_{\text{rabbit}}(t) - E|JX_{\text{rabbit}}(t) - X_m(t)|, \\ \quad \text{if } F(Y) < F(X(t)) \\ Z: Y + S \times \text{LF}(D), \text{ if } F(Z) < F(X(t)) \end{cases} \quad (6)$$

1.2.2 改进的哈里斯鹰优化算法

传统的 HHO 算法与其他群智能优化算法一样, 在求解复杂优化问题时, 存在易陷入局部最优且收敛精度不高等缺点。针对以上问题, 文献[20]引入了指数能量方程、正弦逃逸距离方程、柯西变异和 Circle 混沌扰动对传统的 HHO 算法进行改进, 改进的哈里斯鹰算法记为 THHO。

1) 指数能量方程

逃逸能量 E 的取值决定了种群是进行探索还是局

部开采, 但在传统的 HHO 算法中 E 是通过线性递减进行更新, 即在迭代后期种群只进行局部开采, 这使得算法易陷入局部最优, 因此本文对 E 作如下改进:

$$E_0 = ce^{-\frac{t}{T}} \quad (7)$$

$$E = E_0 \times (2 \times r_6 - 1) \quad (8)$$

式中: c 为常数 2; r_6 为 $[0, 1]$ 之间均匀分布的随机数。

2) 逃逸距离方程

猎物逃逸能量 E 的变化必然会引起其在逃跑过程中跳跃距离 J 的变化, 因此设置 J 的取值如下:

$$J = \begin{cases} \left| \sin(\frac{\pi E}{2}) \right|, & |E| \geq 0.5 \\ E^2, & |E| < 0.5 \end{cases} \quad (9)$$

3) 柯西变异和 Circle 混沌扰动

柯西变异来源于连续型概率分布的柯西分布, 主要特点为零处峰值较小, 从峰值到零值下降缓慢, 可使变异范围更均匀。在种群个体位置更新中引入柯西变异, 提升算法的搜索能力。变异公式为:

$$\text{mutation}X(t) = X(t)(1 + \tan(\pi(u - 0.5))) \quad (10)$$

式中: $\text{mutation}X(t)$ 为个体经过柯西变异后的位置; $X(t)$ 为个体原来位置; u 为区间 $[0, 1]$ 内的随机数。

混沌扰动可增强算法的全局搜索能力。在全局探索和局部开采后, 对种群中适应度值较高的个体进行 Circle 混沌扰动, 与柯西变异同时发挥作用, 以增加种群的多样性, 防止陷入局部最优。

$$\text{Circle}X(t) = X(t) + a - \text{mod}(\frac{b}{2\pi} \cdot \sin(2\pi X(t)), 1) \quad (11)$$

式中: $\text{Circle}X(t)$ 为个体经过混沌扰动后的位置; $X(t)$ 为个体原来位置; a 为常数 0.5, b 为常数 2.2。

1.3 l_p 范数

设 J 表示基于 l_p 范数的误差绝对值之和, 则:

$$J = (\sum_{t=1}^N |e_t|^p)^{\frac{1}{p}} = (\sum_{t=1}^N \left| \sum_{i=1}^n l_i e_{it} \right|^p)^{\frac{1}{p}} \quad (12)$$

式中: e_t 为第 t 时点组合预测的误差; e_{it} 为第 t 时点的第 i 种预测方法的误差; l_i ($\sum_{i=1}^n l_i = 1$) 为第 i 种预测方法在组合预测模型中的权重; p 可取不同值, 本文 p 取 2; N 为模态分量个数; n 为预测方法种类数。

基于 l_p 范数的误差平方绝对值之和能够说明两个序列的接近程度, 其数值越小, 表明预测值越接近实际值, 模型的预测效果越好, 反之则说明两个序列相差较大, 预测效果较差。但是在预测过程中难

免会出现误差,因此构建最优化组合预测模型为:

$$\min J = \left(\sum_{t=1}^N |e_t|^p \right)^{\frac{1}{p}} = \left(\sum_{t=1}^N \left| \sum_{i=1}^n l_i e_{it} \right|^p \right)^{\frac{1}{p}} \quad (13)$$

$$\text{s. t. } \begin{cases} \sum_{i=1}^n l_i = 1 \\ l_i \geq 0, i = 1, 2, \dots, n \end{cases}$$

2 实证结果及分析

2.1 预测结果评估指标

本文选取均方误差 (MSE)、平均绝对误差 (MAE)、均方根误差 (RMSE) 和平均绝对百分比误差 (MAPE) 对预测结果的有效性进行评估。

$$\begin{cases} R_{\text{MSE}} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \\ R_{\text{MAE}} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \\ R_{\text{RMSE}} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \\ R_{\text{MAPE}} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \end{cases} \quad (14)$$

式中: \hat{y}_i 为预测值; y_i 为实际值。

2.2 CEEMDAN_ARIMA_指数平滑预测模型

首先对广东省的碳交易价格序列进行 CEEMDAN 分解,根据分解序列的均值将其重组为高频率序列和低频率序列。将 2019 年 1 月 2 日至 2021 年 2 月 10 日的数据作为训练集,对高频率序列建立 ARIMA 模型,对低频率序列建立指数平滑模型,然后对 2021 年 2 月 18 日至 2021 年 3 月 18 日的 20 条数据进行预测,通过将高频率序列的预测结果和低频率序列的预测结果加和,对广东省碳交易价格进行预测。

2.2.1 模型 I 数据选取

本文选取广东省碳交易每日的收盘价格作为模

型 I 预测研究的基础数据。利用 Python 从碳交易网抓取 2019 年 1 月 2 日至 2021 年 3 月 18 日之间广东省碳交易的相关数据,剔除法定节假日等因素的影响,共包含 536 条数据。

2.2.2 CEEMDAN 分解与重组

由于碳交易价格具有不平稳、非线性特征,因此对 2019 年 1 月 2 日至 2021 年 3 月 18 日之间广东省碳交易价格进行 CEEMDAN 分解,由图 1 可知,CEEMDAN 将原始的碳交易价格数据按照频率的高低分为 7 组不同频率的 IMF 序列及一组残余趋势序列 Res。其本征模函数 (IMF) 和残余项如图 1 所示。

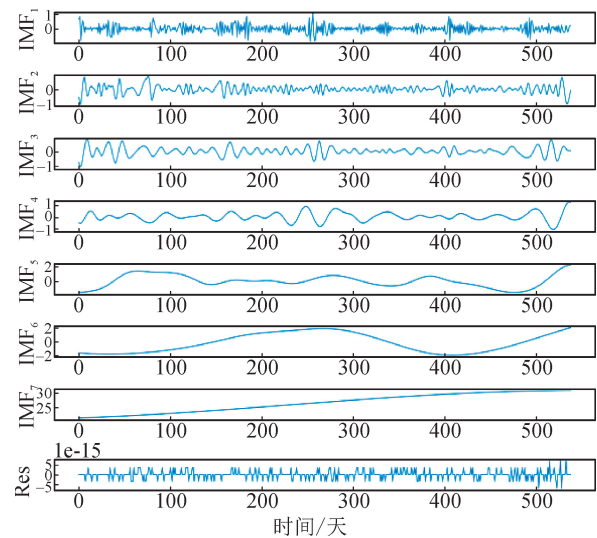


图 1 CEEMDAN 分解产生的 IMF 及 Res

Fig. 1 IMF and Res generated by CEEMDAN decomposition

通过计算本征模函数 IMF 的均值,将广东省的碳交易价格分解序列重构成高频分量和低频分量,残余项的数值忽略不计。由表 1 可知,IMF₁ ~ IMF₆ 的均值与 0 相近,而 IMF₇ 的分量显著偏离 0,因此,本文将 IMF₁ ~ IMF₆ 加和重构为高频分量,将 IMF₇ 重构为低频分量。

表 1 各个 IMF 分量的均值

Tab. 1 Mean value of each IMF component

分量	IMF ₁	IMF ₂	IMF ₃	IMF ₄	IMF ₅	IMF ₆	IMF ₇
均值	0.003 9	-0.002 3	0.003 0	0.022 0	0.008 4	-0.213 8	26.480 1

2.2.3 高频率序列 ARIMA 预测

对高频率序列建立差分整合移动平均自回归 (ARIMA) 模型,为验证该模型的可靠性,选取支持向量机 (SVM) 模型和长短时记忆网络 (LSTM) 模型作为基准对照模型。由于 ARIMA 模型要求序列是平稳序列,因此要对高频率序列进行平稳性检验。

通过 SPSS 对训练集的高频率序列进行自相关 (ACF) 和偏自相关 (PACF) 检验,结果如图 2 所示。

由图 2 可知,高频率序列的 ACF 不趋于 0,说明序列具有很强的长期相关性,因此该序列为非平稳序列。对高频率序列进行一阶差分,使其平稳化,并对一阶差分序列进行平稳性检验,由图 3 可知,高频率序列一阶差分

的 ACF 和 PACF 均为拖尾,因此其为平稳序列。

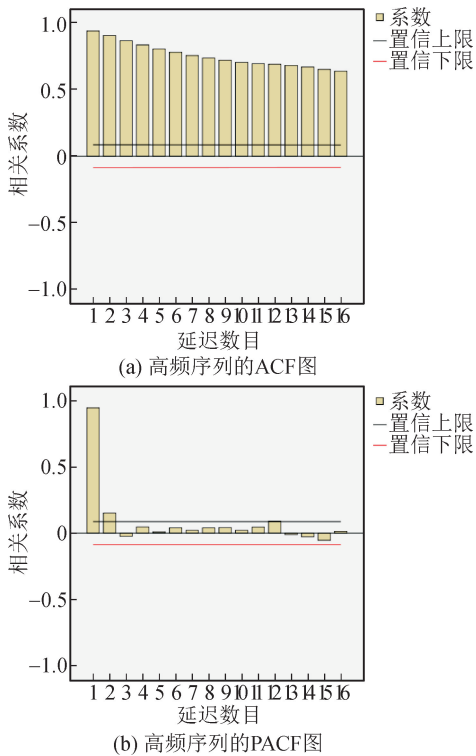


图 2 高频序列的 ACF 和 PACF 图

Fig. 2 ACF and PACF diagram of high frequency sequence

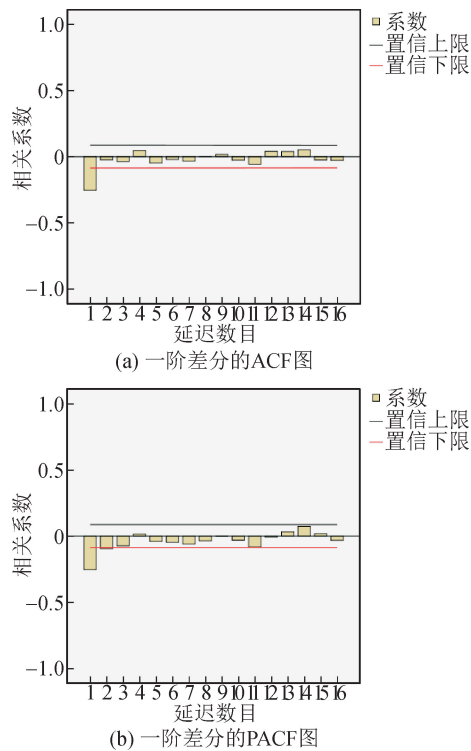


图 3 一阶差分的 ACF 和 PACF 图

Fig. 3 ACF and PACF diagram of first-order difference

由图 3 可知,一阶差分的 ACF 是一个逐渐趋于 0 的拖尾,ACF 自 1 阶过后都落在置信区间内,所以是 MA(1)模型;一阶差分的 PACF 是一个逐渐趋于

0 的拖尾,PACF 自 3 阶过后基本都落在置信区间内,所以是 AR(3)模型,因此应对高频序列建立 ARIMA(3,1,1)模型。

ARIMA(3,1,1)模型、支持向量机模型(SVM)和长短时记忆网络模型(LSTM)对高频序列的预测精度如表 2 所示。

表 2 高频序列预测精度对比
Tab. 2 Comparison of prediction accuracy of high frequency series

模型	R_{MSE}	R_{MAE}	R_{RMSE}	R_{MAPE}
ARIMA	0.256 7	0.340 0	0.506 6	12.445 2
SVM	0.828 1	0.643 0	0.910 0	18.553 3
LSTM	1.685 4	0.999 8	1.298 2	25.816 8

由表 2 可知,ARIMA(3,1,1)模型比支持向量机(SVM)模型的 MSE、MAE、RMSE、MAPE 值分别降低了 69.00%、47.12%、44.33%、32.92%,且 ARIMA(3,1,1)模型的预测精度远高于 LSTM 模型,说明 ARIMA(3,1,1)模型对高频序列的预测效果优于其对照模型。

2.2.4 低频序列指数平滑预测

由于低频序列具有线性、平稳特征,因此对低频序列采用指数平滑预测。通过对低频数据训练集进行不同平滑次数、不同平滑系数的研究及对比分析,可知当平滑次数为 3 次、平滑系数为 0.1 时,模型的预测效果最佳。为验证指数平滑模型对低频序列预测效果的可靠性,选取支持向量机(SVM)模型和长短时记忆网络(LSTM)模型作为基准对照模型,结果如表 3 所示。

表 3 低频序列预测精度对比
Tab. 3 Comparison of prediction accuracy of low frequency series

模型	R_{MSE}	R_{MAE}	R_{RMSE}	R_{MAPE}
指数平滑	1.54E-05	0.003 3	0.003 9	0.011 0
SVM	0.015 1	0.122 2	0.122 7	0.403 3
LSTM	0.275 1	0.524 2	0.524 4	1.729 9

由表 3 可知,当平滑次数为 3 次、平滑系数为 0.1 时,3 次指数平滑模型的 MSE、MAE、RMSE、MAPE 值分别为 1.54E-05、0.003 3、0.003 9、0.011 0,均远低于 SVM 和 LSTM 模型,说明 3 次指数平滑模型对低频序列的预测效果优于 SVM 和 LSTM 模型。

2.2.5 模型 I 预测结果

相比于 SVM 模型和 LSTM 模型,对高频序列

建立 ARIMA 模型、对低频序列建立指数平滑模型预测效果最优。将 ARIMA 高频序列预测结果和指数平滑模型低频序列预测结果加和,作为 2021 年 2 月 18 日至 2021 年 3 月 18 日的碳交易价格的预测值,并由此得出模型 I 的预测精度,结果如表 4 所示。

表 4 模型 I 预测精度
Tab. 4 Prediction accuracy of Model I

R_{MSE}	R_{MAE}	R_{RMSE}	R_{MAPE}
0.255 7	0.340 0	0.505 7	1.008 2

通过对碳交易价格时间序列进行 CEEMDAN 分解、重组,然后对高频序列建立 ARIMA 模型,对低频序列建立指数平滑模型,得到模型 I 预测结果 MSE、MAE、RMSE、MAPE 值分别为 0.255 7、0.340 0、0.505 7、1.008 2。

2.3 THHO_ELM 预测模型

另一方面,综合考虑碳交易价格的行情指标和经济指标两个维度,从中找出与碳交易价格相关的 15 个影响因素,通过 Pearson 相关系数选取与下一日碳交易价格高度相关的 6 个变量并将其作为解释变量,然后将下一日碳交易价格作为被解释变量,建立多策略融合改进的哈里斯鹰优化极限学习机模型 (THHO_ELM),从而对广东省碳交易价格进行预测。

2.3.1 模型 II 数据选取

利用 Python 从碳交易网抓取广东省碳交易的相关数据,选取广东省碳交易的最高价、最低价、开盘价、收盘价和交易额 5 个变量作为行情指标,通过多日内碳交易的行情指标数据计算出碳交易的 10 个经济指标数据,其中包括 12 日短期移动平均线 (MA12)、26 日长期移动平均线 (MA26)、平滑异同移动平均线 (MACD)、简易波动指标 (EMV)、随机指标 (K、D、J)、相对强弱指标 (RSI)、乖离率和心理线 (PSY)。根据当日碳交易的行情指标、经济指标与下一日收盘价的 Pearson 系数来判定二者之间的相关性强弱,结果如表 5 所示。

由表 5 可知,碳交易的最高价、最低价、开盘价、收盘价、MA12 和 MA26 与下一日收盘价高度相关,其相关系数分别为 0.971、0.962、0.984、0.989、0.982、0.974。因此,论文将 2019 年 1 月 2 日至 2021 年 2 月 10 日的数据作为训练集,将广东省碳交易的最高价、最低价、开盘价、收盘价、MA12 和 MA26 作为解释变量,下一日碳交易收盘价作为被解释变量,建立哈里斯鹰优化极限学习机模型 (THHO_ELM)对 2021 年 2 月 18 日至 2021 年 3

月 18 日的碳交易收盘价进行预测。

表 5 碳交易价格的 Pearson 相关系数
Tab. 5 Pearson correlation coefficient of carbon trading price

指标类别	指标名称	Pearson 相关系数
行情指标	最高价	0.971
	最低价	0.962
	开盘价	0.984
	收盘价	0.989
	交易额	0.057
经济指标	MA12	0.982
	MA26	0.974
	MACD	0.047
	EMV	-0.191
	K	-0.139
	D	-0.295
	J	0.199
	RSI	-0.047
	乖离率	-0.019
	PSY	-0.060

2.3.2 THHO_ELM 算法流程

为了提高传统 ELM 算法的预测准确度,论文利用改进的哈里斯鹰算法对 ELM 算法的权值和阈值进行优化。其中种群初始规模为 50,最大迭代次数为 500,ELM 算法的权值取值范围为 $[-1,1]$,ELM 算法的阈值取值范围为 $[-10,10]$,THHO_ELM 算法具体流程如下:

- 1) 确定 ELM 模型的结构,其中主要包括输入层神经元个数、隐含层神经元个数、输入权值数、隐含层节点阈值数;
- 2) 根据 ELM 模型进行哈里斯鹰种群初始化,设置种群规模、最大迭代次数等参数,在给定范围内随机产生哈里斯鹰种群;
- 3) 采用训练集、初始权值和阈值训练 ELM 模型并对测试集进行预测,将测试集的 MSE 值作为个体的适应度值,计算哈里斯鹰种群的初始适应度值;
- 4) 根据逃逸能量 E 和随机数 r 的值选择相应的阶段或围攻策略,根据位置更新公式对哈里斯鹰种群进行更新;
- 5) 计算每个个体的适应度值,更新全局最优值;
- 6) 判断是否达到算法终止条件,若达到条件,则输出全局的最优适应度值,否则循环执行步骤 4)

和 5);

7) 将哈里斯鹰算法得到的最优位置参数赋值给 ELM 模型的输入权值和隐层节点的阈值。在训练集样本下,计算 ELM 的输出权值矩阵,并对测试集样本进行预测。

2.3.3 模型 II 预测结果

将每日的碳交易最高价、最低价、开盘价、收盘价、MA12 和 MA26 作为解释变量,下一日收盘价作为被解释变量,建立 ELM 模型对碳交易价格进行预测。为了验证 ELM 模型预测效果的可靠性,选取随机森林模型(RF)、支持向量回归机模型(SVR)进行对比试验,结果如表 6 所示。

表 6 ELM 模型预测精度
Tab.6 ELM model prediction accuracy

模型	R_{MSE}	R_{MAE}	R_{RMSE}	R_{MAPE}
RF	4.459 1	1.516 6	2.111 6	4.346 1
SVR	1.261 2	0.833 7	1.123 0	2.461 6
ELM	0.270 2	0.380 1	0.519 8	1.128 2

由表 6 可知,ELM 模型的预测效果要优于 RF 模型和 SVR 模型,因此本文选择利用群智能优化算法对 ELM 模型做进一步优化。

本文选用粒子群算法(PSO)、灰狼优化算法(GWO)和 HHO 算法优化 ELM 模型,对碳交易价格进行预测。结果如表 7 所示。

表 7 ELM 优化模型预测精度
Tab.7 ELM optimization model prediction accuracy

模型	R_{MSE}	R_{MAE}	R_{RMSE}	R_{MAPE}
PSO_ELM	0.149 8	0.307 2	0.387 0	0.920 5
GWO_ELM	0.245 2	0.397 3	0.495 2	1.171 2
HHO_ELM	0.142 7	0.282 9	0.377 7	0.839 9

由表 7 可知,PSO 算法、GWO 算法和 HHO 算法都对 ELM 模型具有一定的优化效果,且 HHO 算法的 MSE、MAE、RMSE、MAPE 值均低于 PSO 算法和 GWO 算法,即 HHO 算法对 ELM 模型的优化效果优于 PSO 算法和 GWO 算法。因此选用 HHO 算法对 ELM 模型做进一步优化。

THHO_ELM 模型 的 MSE、MAE、RMSE、MAPE 值分别为 0.095 3、0.263 1、0.308 6、0.774 2。由 ELM 模型、HHO_ELM 模型和 THHO_ELM 模型 的 MSE、MAE、RMSE、MAPE 值可知,HHO 算法和 THHO 算法都能提升 ELM 模型的预测精度,且 THHO_ELM 模型的预测精度优于 HHO_ELM 模型。

为了综合模型优势,将 CEEMDAN-ARIMA-指数平滑模型和 THHO_ELM 模型进行组合,建立基于 l_p 范数的误差平方绝对值模型。

2.4 基于 l_p 范数的组合预测模型

在 CEEMDAN-ARIMA-指数平滑模型和 THHO_ELM 模型的基础上进行 l_p 范数组组合预测。当 CEEMDAN-ARIMA-指数平滑模型的权值为 0.211、THHO_ELM 模型的权值为 0.789 时,模型的组合预测误差最小。基于 l_p 范数组组合预测模型及单一模型的预测结果如表 8 所示,对比图如图 4 所示。

表 8 l_p 范数组组合预测模型及单一模型的预测结果
Tab.8 Prediction results of single model and combined forecasting model based on l_p norm

模型	R_{MSE}	R_{MAE}	R_{RMSE}	R_{MAPE}
LSTM 模型 (加入网格搜索)	0.976 5	1.335 8	0.035 3	
LGBM 模型 (加入网格搜索)	0.870 1	1.171 3	0.029 2	
LSTM-LSTM-LGBM 模型(加入网格搜索)	0.787 2	1.128 1	0.027 9	
CEEMDAN-ARIMA- 指数平滑模型	0.255 7	0.340 0	0.505 7	1.008 2
THHO_ELM 模型	0.095 3	0.263 1	0.308 6	0.774 2
l_p 范数组组合预测模型	0.090 8	0.238 3	0.301 4	0.698 8

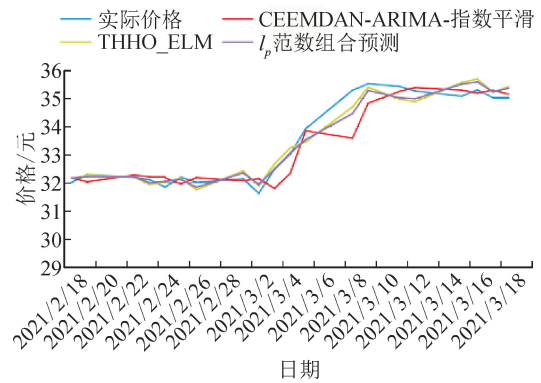


图 4 各模型预测对比图

Fig.4 Comparison of prediction of each model

由表 8 可知,基于 l_p 范数组组合预测模型相比于最优的单一预测模型(THHO_ELM),其 MSE、MAE、RMSE、MAPE 值分别降低了 4.72%、9.43%、2.33%、9.74%,说明基于 l_p 范数组组合预测模型优于 CEEMDAN-ARIMA-指数平滑模型和 THHO_ELM 模型。为进一步验证基于 l_p 范数组组合预测模型的预测效果,本文选取文献[21]作为

对照,结果表明,基于 l_p 范数的组合预测模型优于对照模型。

此外,考虑到不同碳交易试点的差异性,论文又在我国的东、中、西部分别选取上海、湖北和重庆碳交易试点,并利用上述模型进行预测,结果如表 9 所示。

表 9 碳交易试点模型对比

Tab. 9 Comparison of carbon trading pilot models

试点	模型	R_{MSE}	R_{MAE}	R_{RMSE}	R_{MAPE}
上海	CEEMDAN-ARIMA- 指数平滑模型	0.091 4	0.273 0	0.302 4	0.654 2
	THHO_ELM 模型	0.093 6	0.164 7	0.306 0	0.387 4
	l_p 组合预测模型	0.062 8	0.115 9	0.250 6	0.243 6
湖北	CEEMDAN-ARIMA- 指数平滑模型	2.939 6	1.435 8	1.714 5	23.862 1
	THHO_ELM 模型	1.435 8	0.951 3	1.198 2	5.851 3
	l_p 组合预测模型	1.131 6	0.877 8	1.063 7	2.856 2
重庆	CEEMDAN-ARIMA- 指数平滑模型	8.520 8	2.257 3	2.919 0	8.823 6
	THHO_ELM 模型	0.095 2	0.263 1	0.308 6	0.775 2
	l_p 组合预测模型	0.094 7	0.260 0	0.307 8	0.764 1

由表 9 可以看出,基于 l_p 范数的组合预测模型优于其单一预测模型。因此认为,该组合模型对碳交易价格有较好的预测效果。

3 结 语

论文在传统的碳交易价格预测模型的基础上综合考虑了碳交易价格本身和碳交易价格的相关影响因素,分别建立了 CEEMDAN-ARIMA-指数平滑模型和改进的哈里斯鹰优化极限学习机的碳价格预测模型。为综合二者优势,论文建立了基于 l_p 范数的组合预测模型,实证结果表明,基于 l_p 范数的组合预测模型相比于最优的单一预测模型,其 MSE、MAE、RMSE、MAPE 值分别降低了 4.72%、9.43%、2.33%、9.74%,取得了较好的预测效果。该组合预测模型可为预测碳交易价格提供参考和借鉴,同时该模型也可应用于其它行业的预测。

参考文献:

[1] 胡迁林,赵明. “十四五”时期现代煤化工发展思考[J]. 中国煤炭,2021,47(3):2-8.
HU Qianlin, ZHAO Ming. Thinking on the development of modern coal chemical industry during the 14th Five-Year Plan period [J]. China Coal, 2021, 47(3): 2-8.

[2] 姚星,陈灵杉,张永忠. 碳交易机制与企业绿色创新:基于三重差分模型[J]. 科研管理,2022,43(6):43-52.
YAO Xing, CHEN Lingshan, ZHANG Yongzhong. The carbon trading mechanism and corporate green innovation in China: a study based on the DDD model[J]. Science Research Management, 2022, 43(6): 43-52.

[3] 鞠斐,袁野. 碳价格对中国“双碳”目标实现的影响与策略研究[J]. 价格月刊,2022(7):19-24.
JU Fei, YUAN Ye. Research on the influence and strategy of carbon price on the realization of China's "dual carbon" goals[J]. Prices Monthly, 2022(7): 19-24.

[4] LI Guohui, NING Zhiyuan, YANG Hong, et al. A new carbon price prediction model[J]. Energy, 2021, 239(3):122324.

[5] SUN Guoqiang, CHEN Tong, WEI Zhinong, et al. A carbon price forecasting model based on variational mode decomposition and spiking neural networks[J]. Energies, 2016, 9(1): 1-16.

[6] QIN Quande, HE Huangda, LI Li, et al. A novel decomposition-ensemble based carbon price forecasting model integrated with local polynomial prediction[J]. Computational Economics, 2020, 55(4): 1249-1273.

[7] 崔金鑫,邹辉文. 基于 CEEMDAN-MR-PE-NLE 多频优化组合模型的碳金融市场价格的预测[J]. 数学的实践与认识,2020,50(3):105-120.
CUI Jinxin, ZOU Huiwen. Carbon financial market price forecasting based on CEEMDAN-MR-PE-NLE multi-frequency optimization combined model [J]. Mathematics in Practice and Theory, 2020, 50(3): 105-120.

[8] 姚奕,吕静,章成果. 湖北碳市场价格形成机制及价格预测[J]. 统计与决策,2017(19):166-169.

[9] ZHANG Fang, WEN Nuan. Carbon price forecasting: a novel deep learning approach[J]. Environmental Science and Pollution Research, 2022, 29:54782-54795.

[10] XU Hua, WANG Minggang, JIANG Shumin, et al. Carbon price forecasting with complex network and extreme learning machine[J]. Physica A: Statistical Mechanics and its Applications, 2020, 545:122830.

[11] YANG Yi, GUO Honggang, JIN Yu, et al. An ensemble prediction system based on artificial neural networks and deep learning methods for deterministic and probabilistic carbon price forecasting[J]. Frontiers in Environmental Science, 2021(9).

[12] 彭武元,陈思宇. 中国碳排放试点市场碳交易价格分析及预测[J]. 技术经济,2020,39(3):102-110.
PENG Wuyuan, CHEN Siyu. Analysis and forecast of carbon trading price in China's carbon emission pilot market [J]. Journal of Technology Economics,

- 2020, 39(3): 102-110.
- [13] ZHANG Lu, ZHANG Junbiao, XIONG Tao, et al. Interval forecasting of carbon futures prices using a novel hybrid approach with exogenous variables[J]. *Discrete Dynamics in Nature and Society*, 2017: 1-12.
- [14] 蒋锋,彭紫君. 基于混沌 PSO 优化 BP 神经网络的碳价预测[J]. *统计与信息论坛*, 2018, 33(5): 93-98.
JIANG Feng, PENG Zijun. Forecasting of carbon price based on BP neural network optimized by chaotic PSO algorithm[J]. *Journal of Statistics and Information*, 2018, 33(5): 93-98.
- [15] XIE Qiwei, HAO Jingjing, LI Jingyu, et al. Carbon price prediction considering climate change: a text-based framework[J]. *Economic Analysis and Policy*, 2022, 74(C): 382-401.
- [16] 刘金培,郭艺,陈华友,等. 基于非结构数据流行学习的碳价格多尺度组合预测[J]. *控制与决策*, 2019, 34(2): 279-286.
LIU Jinpei, GUO Yi, CHEN Huayou, et al. Multi-scale combined forecast of carbon price based on manifold learning of unstructured data[J]. *Control and Decision*, 2019, 34(2): 279-286.
- [17] 魏宇,张佳豪,陈晓丹. 基于 DMS 和 DMA 的我国碳排放权交易价格预测方法——来自湖北碳市场的经验证据[J]. *系统工程*, 2022, 40(4): 1-16.
WEI Yu, ZHANG Jiahao, CHEN Xiaodan. Forecasting China's carbon trading price in a DMS and DMA framework——evidence from the Hubei carbon market[J]. *Systems Engineering*, 2022, 40(4): 1-16.
- [18] 岳有军,刘英翰,赵辉,等. 基于 CEEMDAN-SE 和 DBN 的短期电力负荷预测[J]. *电测与仪表*, 2020, 57(17): 59-65.
YUE Youjun, LIU Yinghan, ZHAO Hui, et al. Short-term load forecasting based on CEEMDAN-SE and DBN[J]. *Electrical Measurement & Instrumentation*, 2020, 57(17): 59-65.
- [19] 郭雨鑫,刘升,高文欣,等. 精英反向学习与黄金正弦优化的 HHO 算法[J]. *计算机工程与应用*, 2022, 58(10): 153-161.
GUO Yuxin, LIU Sheng, GAO Wenxin, et al. Elite opposition-based learning golden-sine Harris Hawks optimization[J]. *Computer Engineering and Applications*, 2022, 58(10): 153-161.
- [20] 赵峰,徐丹华. 基于混沌扰动与柯西变异的 HHO 算法[J/OL]. *安徽大学学报(自然科学版)*, 2022: 1-12 [2023-06-14]. <http://kns.cnki.net/kcms/detail/34.1063.N.20220829.1822.021.html>.
ZHAO Feng, XU Danhua. HHO algorithm based on chaotic disturbance and Cauchy mutation [J/OL]. *Journal of Anhui University (Natural Science Edition)*, 2022: 1-12 [2023-06-14]. <http://kns.cnki.net/kcms/detail/34.1063.N.20220829.1822.021.html>.
- [21] 呼雪芳. 基于 CEEMDAN 分解的碳交易价格组合预测模型研究[D]. 广州:华南理工大学, 2021.
HU Xuefang. Research on the prediction model of carbon trading price portfolio based on CEEMDAN decomposition[D]. Guangzhou: South China University of Technology, 2021.

(责任编辑 周 蓓)