

基于注意力机制组合模型的全国碳价分析及预测

豪 谕, 谷爱玲, 张荣森

广东工业大学数学与统计学院, 广东 广州

收稿日期: 2025年6月4日; 录用日期: 2025年6月27日; 发布日期: 2025年7月3日

摘 要

碳排放权交易价格是碳交易市场的核心要素, 为了帮助企业、投资者和政府优化碳市场参与行为, 需要对碳排放权交易价格进行合理有效的预测。本文利用皮尔逊相关系数法(PCC)提取碳价关键影响因素, 然后运用由Atrous Spatial Pyramid Pooling (ASPP)、以Sophia为优化器的LSTM模型和集成学习XGBoost模型组合而成, 以及添加了基于注意力机制Efficient Multi-Scale Attention (EMA)的组合模型, 对全国碳市场交易价格进行预测, 并与单一模型对比, 通过模型预测值的MSE值、RMSE值、MAE值和R²值对比预测精度, 检验组合模型的有效性。对比结果表明: 注意力机制组合模型的预测精度最高, 是一种有效的且精度高的碳价预测模型。

关键词

皮尔逊相关系数, 注意力机制, 组合模型

National Carbon Price Analysis and Forecasting Based on Attention Mechanism Portfolio Modeling

Yu Hao, Ailing Gu, Rongsen Zhang

School of Mathematics and Statistics, Guangdong University of Technology, Guangzhou Guangdong

Received: Jun. 4th, 2025; accepted: Jun. 27th, 2025; published: Jul. 3rd, 2025

Abstract

Carbon emission rights trading price is the core element of carbon trading market, in order to help

文章引用: 豪谕, 谷爱玲, 张荣森. 基于注意力机制组合模型的全国碳价分析及预测[J]. 数据挖掘, 2025, 15(3): 213-227. DOI: 10.12677/hjdm.2025.153018

enterprises, investors and governments optimize the carbon market participation behavior, it is necessary to make reasonable and effective prediction of carbon emission rights trading price. In this paper, the Pearson Correlation Coefficient (PCC) method is used to extract the key influencing factors of carbon price, and then a combination of Atrous Spatial Pyramid Pooling (ASPP), LSTM model with Sophia as the optimizer, and Integrated Learning XGBoost model is applied, as well as the addition of an attention-based mechanism Efficient Multi-Scale Attention (EMA) combined model to predict the trading price in the national carbon market, and compared with a single model to test the effectiveness of the combined model by comparing the prediction accuracy with the MSE, RMSE, MAE and R^2 values of the model prediction values. The comparison results show that the combined model of attention mechanism has the highest prediction accuracy and is an effective and highly accurate carbon price prediction model.

Keywords

Pearson Correlation Coefficient, Attention Mechanism, Combined Model

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

在全球气温逐渐变暖的环境下, 减排已成为世界各国重点关注的核心问题。我国在实现减排问题上十分重视, 2020 年 9 月, 习主席在联合国大会上发布了碳达峰和碳中和的“双碳”目标。碳排放权交易价格(简称“碳价”)是碳市场的核心要素, 对碳价的准确预测, 可以帮助企业和投资者规避风险, 也可以帮助政府更好制定有效措施, 使市场机制更好发挥减排作用。本文基于注意力机制组合模型预测全国碳排放权交易市场碳价的变化趋势, 为政府和相关企业提供决策支持。

目前, 国内外学者正积极从多个角度针对碳价的影响因素和预测模型进行相关研究。关于影响碳价的各种因素, Bredin 和 Muckley [1]利用 2005 年到 2009 年的欧盟碳价格分析了经济增长、能源价格和气候条件对碳价的影响。Piia 等[2]用股票指数作为描述宏观经济形势的指标, 发现碳价受宏观经济因素影响显著。张玲和曹峰[3]探讨了环境规制水平和地理距离对碳价的影响。姜瑜和吴哲宇[4]利用多元回归模型检验得出, 金融市场、能源价格、空气质量、国际碳市场、技术进步等指标均与碳交易价格显著相关。王小燕等[5]针对广州碳市场碳价, 用国际碳价、国内外经济指标、国外能源指标、国内能源指标、气候环境和宏观政策构建指标体系研究碳价影响因素。吴慧娟和张智光[6]研究了碳价的时间波动及区域差异与政策、经济与环保因素的关系。

关于碳价预测模型, 现有碳价预测模型主要分两类: 基于碳价历史数据的时间序列模型和基于碳价影响因素的机器学习模型。在时间序列分析模型方面, Zhang 和 Xu 等[7]利用 GARCH 模型研究了深圳的碳交易市场, 预测其碳价波动。Jiang 和 Wu [8]将碳价原始数据序列进行了分解, 并采用 ARIMA-RF 集成模型对分解的碳价序列进行预测。王娜[9]运用 Boosting-ARMA 时间序列算法对欧盟碳期货价格进行预测。Zhu 等[10]运用 EMD-ARIMA-ADD 模型对欧盟碳交易市场碳价实现了精准预测。在神经网络模型方面, 金林等[11]用 Lasso 回归法对影响碳价变量开展筛选, 并采用灰色 BP 神经网络预测碳价。Li 等[12]以湖北碳市场碳价为实验目标, 并选取三种能源价格为关键影响因素, 采用 LSTM 模型对碳价进行预测。朱亮亮等[13]以全国碳市场碳价为研究对象, 运用 ARIMA 模型、多项式回归算法和 LSTM 模型进行了预测, 结果表明皮尔逊相关系数分析(PCC)-LSTM 模型的精度相对较高。呼雪芳[14]对广东碳价序列运用

CEEMDAN 算法进行分解, 并通过 LSTM-LSTM-LGBM 组合预测模型进行预测。高长征等[15]以湖北碳市场为研究对象, 运用 CEEMDAN 算法对碳价原始序列进行分解, 并通过 Transformer 模型实现了对碳价的预测。Wang 等人[16]为克服风能频繁的局部波动, 集成了自回归误差校正模型, 进一步提升了预测的准确性。Xiaolu Du 等[17]基于 Lmdi 分解和计量模型分析研究能源价格对碳排放权需求量的影响, 并提出三点建议。Liu 等[18]结合了时间卷积网络和注意力机制, 利用双通道注意力模块来增强 TCN 处理序列内部依赖性的能力, 提升了模型的预测精度。

通过上述文献可以发现, 现在国内外在碳价预测领域已取得了相当的成果, 但仍然有可以完善的地方: 一方面, 目前已有的成果大多以欧盟碳市场和我国的试点碳市场为研究对象, 针对全国碳市场碳价预测的研究较少; 另一方面, 预测方法多为单一的预测模型, 多种模型组合而成的组合预测模型较少。为了使研究对象更具有一般性, 以及开拓更有效的预测, 在考虑到碳价序列同时具有非线性和线性的特征的基础上, 本文选取全国碳交易市场碳价作为研究对象, 首先, 通过皮尔逊相关系数法(PCC)确定碳价关键影响因素, 而后通过结合多种先进技术, 提升 LSTM 模型在复杂市场中预测价格的能力。LSTM 被选用是因为其在处理时间序列数据上的优势, 能够捕捉价格波动中的长期依赖关系。而加入 Efficient Multi-Scale Attention (EMA), 可以让模型在多尺度上关注重要的市场信号, 从短期趋势到长期变化, 提高预测的准确性。Atrous Spatial Pyramid Pooling (ASPP)的引入, 扩展了感受野, 能够融合来自不同尺度的上下文信息, 从而更全面地理解市场动态。使用 Sophia 优化器, 则是为了加速训练过程, 提升收敛速度, 避免过拟合, 并确保模型能够高效学习到有价值的特征。最后, 通过 XGBoost 微调, 进一步提升 LSTM 模型性能, 提升整体的预测精度。

2. 碳价影响因素分析

我国碳价的影响因素主要从经济形势、金融市场、国际碳市场、化石能源价格、气候环境、其他因素这 6 大类指标分析[19]。

经济形势方面, 本文选择沪深 300 指数和标准普尔 500 指数作为评价国内外经济形势的指标[20]。金融市场方面, 本文选取 Shibor 和欧元、美元汇率作为评价金融市场的指标[21]。国际碳市场方面[22], 本文选取欧盟 EUA 碳交易期货价格作为评价国际碳市场的指标。化石能源价格[23], 本文选取动力煤收盘价、大庆原油现货价、LNG 液化天然气市场价作为评价化石能源价格的指标。气候环境方面[24], 本文选取全国共 31 个省会城市及直辖市(除港澳台)的日平均气温和日平均空气质量指数作为评价气候环境的指标。其他因素方面, 本文选取中证新能源指数作为评价新能源的指标以及“低碳”、“碳排放”和“碳交易”搜索指数作为评价搜索数据的指标。

所有数据均来自 Wind 数据库、Choice 数据库、天气网和百度搜索指数。由于选取的数据来自不同网站, 数据可能存在缺失, 本文对缺失数据进行插值来补全交易数据完整性[25]。具体选取碳价影响因素见表 1。

考虑到不同影响因素对碳价会有不同的影响程度[26], 因此除了对碳价影响因素进行分析外, 本文采用皮尔逊相关系数法对表 1 中所选取的影响因素进行分析, 并筛选出影响程度较高的几个变量作为预测的指标。

皮尔逊系数法可以度量两个变量之间的相关程度[27], 该方法一般用来挑选回归模型的输入变量, 因此可以用该方法对关键影响因素进行筛选。被解释变量选取全国碳市场每日碳排放权交易价格的收盘价, 研究收集了从 2021 年 7 月 16 日到 2023 年 9 月 1 日期间除去节假日外共 518 条全国碳市场成交日的价格数据。价格走势见图 1。

采用皮尔逊相关系数法对前面提到的 6 类, 共 15 个碳价影响因素进行相关性分析, 结果见表 2。

Table 1. Carbon price influences
表 1. 碳价影响因素

类别	变量	数据来源
经济形势	沪深 300 指数	Wind 数据库
	标准普尔 500 指数	Wind 数据库
金融市场	Shibor	Wind 数据库
	欧元汇率	Wind 数据库
	美元汇率	Wind 数据库
国际碳市场	EUA 期货价格	Wind 数据库
化石能源价格	动力煤收盘价	Choice 数据库
	大庆原油现货价	Choice 数据库
	LNG 液化天然气市场价	Wind 数据库
气候环境	日平均气温	天气网
	日平均空气质量指数	天气网
其他因素	中证新能源指数	Wind 数据库
	“低碳”搜索指数	百度搜索指数
	“碳排放”搜索指数	百度搜索指数
	“碳交易”搜索指数	百度搜索指数

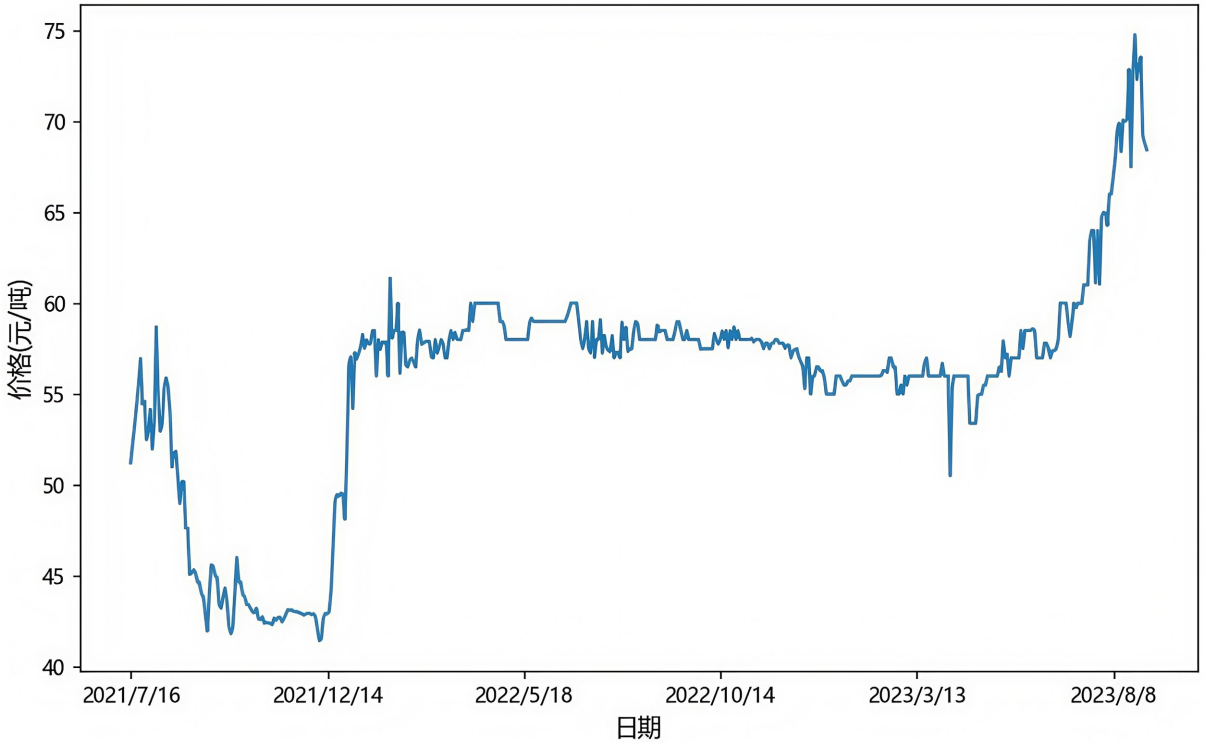


Figure 1. National carbon market closing price
图 1. 全国碳市场收盘价

数据表明，沪深 300 指数、美元汇率、EUA 期货价格和中证新能源指数跟碳价呈现显著相关关系；标准普尔 500 指数、Shibor、大庆原油现货价、LNG 液化天然气市场价、“碳排放”搜索指数跟碳价呈现低度相关关系；欧元汇率、动力煤收盘价、日平均气温、日平均空气质量指数、“低碳”搜索指数和

Table 2. Correlation analysis of influencing factors**表 2.** 影响因素相关性分析

影响因素	相关系数值	相关程度
沪深 300 指数	-0.704	显著负相关
标准普尔 500 指数	-0.34	低度负相关
续 Shibor	-0.393	低度负相关
欧元汇率	0.013	微弱正相关
美元汇率	0.556	显著正相关
EUA 期货价格	0.526	显著正相关
动力煤收盘价	-0.029	微弱负相关
大庆原油现货价	0.417	低度正相关
LNG 液化天然气市场价	-0.318	低度负相关
日平均气温	0.259	微弱正相关
日平均空气质量指数	0.138	微弱正相关
中证新能源指数	-0.697	显著负相关
低碳搜索指数	-0.034	微弱负相关
碳排放搜索指数	-0.45	低度负相关
碳交易搜索指数	-0.29	微弱负相关

“碳交易”搜索指数跟碳价只有微弱相关关系。因此，本文选取沪深 300 指数、美元汇率、EUA 期货价格和中证新能源指数这 4 个具有显著相关性的因素，然后加上碳价的开盘价、收盘价、最高价、最低价和成交量 5 个碳价市场基本特征共 9 个变量作为后续预测碳价模型的关键输入变量。

3. 模型概述

3.1. Sophia 模型

为了增加模型训练时的稳定性，提高碳价预测精度，我们使用 Sophia [28]来优化 LSTM 模型[29]训练时的参数。鉴于神经网络模型训练的成本，优化算法的微小的改进也能大幅减少训练所需的时间和成本。近些年来，Adam [30]及其变体一直是优化算法的佼佼者，而 Adam 复杂的二阶优化器训练成本过大。在本文中，我们使用 Sophia，一种二阶裁剪随机优化算法，这是一种简单且可扩展的二阶优化器，它使用对角海森矩阵的轻量级估计作为预条件器。仅需极低开销即可实现参数维度的曲率自适应学习率缩放；其次，通过“梯度移动平均/海森估计+逐元素裁剪”的更新机制，在保留二阶信息的同时限制单步更新幅度，抑制非凸损失函数中的梯度突变与海森矩阵剧烈波动；最后，采用稀疏海森估计策略，仅间隔性更新二阶信息。Sophia 每隔若干次迭代才估计一次对角海森矩阵，其平均每步的时间和内存开销可以忽略不计。

Sophia 比 Adam 更能适应异质曲率，比牛顿法[31]更能抵抗非凸性和海森矩阵的快速变化，并且还使用了低成本的预条件器。我们用 θ_t 表示时间步长 t 时的参数。在每一步中，我们从数据分布中抽取了一个小的批次，并计算这个批次的损失，记作 $L_t(\theta_t)$ ，用 g_t 表示的梯度，即 $g_t = \nabla L_t(\theta_t)$ ，令 m_t 为梯度的指数移动平均值，计算如下：

$$m_t \leftarrow \beta_1 m_{t-1} + (1 - \beta_1) g_t$$

Sophia 采用基于对角海森矩阵的预条件器，该预条件器会根据不同参数维度的曲率直接调整其更新大小，为减少开销，我们每隔 k 步($k = 10$)才估计一次海森矩阵。在时间步 t 且 $t \bmod k = 1$ (即 t 除以 k 的余数为 1)时，估计器返回一个估计值 h 。与小批量损失函数的梯度类似，估计的对角海森矩阵也可能存在较大的噪声。受 Adam 中指数移动平均(EMA) [32] 的启发，我们也在每次迭代中使用 EMA 对对角海森矩阵的估计值进行去噪。我们每 k 步更新一次 EMA，从而得到对角海森矩阵估计值的如下更新规则：

$$h_t = \beta_2 h_{t-k} + (1 - \beta_2) \hat{h}_t \text{ if } t \bmod k = 1; \text{ else } h_t = h_{t-1}$$

海森矩阵估计的不准确性和其沿轨迹的变化会使二阶信息不可靠。为此，我们仅考虑对角海森矩阵的正元素，并且在更新中引入逐坐标裁剪。对于裁剪阈值 $\rho > 0$ ，设裁剪函数为 $\text{clip}(z, \rho) = \max\{\min\{z, \rho\}, -\rho\}$ ，其中所有操作均按坐标分别进行。更新规则写为：

$$\theta_{t+1} \leftarrow \theta_t - \eta_t \cdot \text{clip}(m_t / \{\gamma \cdot h_t, \varepsilon\}, 1)$$

其中， ε 是一个非常小的常数，用于避免除以 0。且因为

$\eta_t \cdot \text{clip}(m_t / \max\{\gamma \cdot h_t, \varepsilon\}, 1) = (\eta_t / \gamma) \cdot \text{clip}(m_t / \max\{h_t, \varepsilon / \gamma\}, \gamma)$ ，所以我们实质上是按坐标对原始更新 m_t / h_t 进行 γ 范围内的裁剪，然后通过 γ 因子重新调整最终的更新大小。这种重新调整使得更新的规模对 γ 的依赖性降低，因为现在 γ 只控制裁剪项的比例，而所有裁剪项最终都会在更新中被设置为 0。

Sophia 优化器综合了对角 Hessian 预条件器、EMA 以及逐坐标裁剪的策略，既能灵活适应不同参数维度的异质曲率，又在非凸优化中对 Hessian 噪声和快速变化保持较高鲁棒性。它通过周期性地估计对角 Hessian 并使用 EMA 去噪，确保二阶信息更稳定；同时，采用裁剪操作将每个坐标的更新限制在一定范围内，并通过重参数化降低超参数 γ 对更新尺度的直接影响，从而在降低计算开销的同时，有效防止异常更新，提升了碳价预测模型整体训练过程的稳定性和效率。我们在下表中给出了 Sophia 的伪代码如下：

Algorithm: Sophia

Input: θ_1 , learning rate $\{\eta_t\}_{t=1}^T$, hyperparameters $\lambda, \gamma, \beta_1, \beta_2, \varepsilon$, and estimator choice Estimator $\in \{\text{Hutchinson, Gauss-Newton-Bartlett}\}$
 Set $m_0 = 0, v_0 = 0, h_{1-k} = 0$
 for $t = 1$ to T do
 Compute minibatch loss $L_t(\theta_t)$.
 Compute $g_t = \nabla L_t(\theta_t)$.
 $m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t$
 If $t \bmod k = 1$ then
 Compute $\hat{h}_t = \text{Estimator}(\theta_t)$
 $h_t = \beta_2 h_{t-k} + (1 - \beta_2) \hat{h}_t$
 else
 $h_t = h_{t-1}$
 $\theta_t = \theta_t - \eta_t \lambda \theta_t$ (weight decay)
 $\theta_{t+1} = \theta_t - \eta_t \cdot \text{clip}(m_t / \max\{\gamma \cdot h_t, \varepsilon\}, 1)$
 return θ_t

3.2. Efficient Multi-Scale Attention (EMA)

由于碳价格预测受到多种因素的复杂相互作用的影响,为了提高预测碳价格趋势的准确性,在模型中引入注意力是必要的一步[33]。深度学习技术中,我们可以使用注意力机制增强网络对特征的提取能力,引导网络关注重点目标[34]。注意力机制已广泛应用于神经网络,显著提升了其性能。当前常见的注意力机制主要分为三种:通道注意力、空间注意力和混合注意力机制[35]。作为通道注意力的代表, Selective Attention (SE) [36]明确地对跨维度交互进行了建模,以提取通道级注意力。而后续的 Convolutional Block Attention Module (CBAM) [37]在特征图的空间和通道维度之间建立了跨通道和跨空间信息的语义相互依赖关系。本文选用具有更高效的多尺度特征融合能力的 EMA 注意力,且因为并行子结构有助于网络出现过多的顺序处理和过大的深度,鉴于 Coordinate Attention (CA) [38]所使用的并行处理操作, Efficient Multi-Scale Attention (EMA)模块[39]中采用了该策略。EMA 的具体结构和思想如下:

对于输入时给定的碳价数据 $X \in R^{B \times C \times T}$, EMA 将其分解为 G 个子特征,以学习不同语义特征,增加神经元感受野以收集多尺度空间信息[40]。EMA 通过三条平行路线提取分组特征图的注意力权重描述符。由于碳价数据的维度与图像不同,EMA 的两条并行路由位于 1×1 的一维卷积分支中,第三条路由位于 3×3 的一维卷积分支中。在每个分组内,首先通过一维的自适应平均池化操作沿时间维度提取全局特征描述,用于捕获跨通道的全局信息[38]。随后, 1×1 分支对分组特征分别利用核大小为 1×1 一维卷积进行局部跨通道信息交互,并将卷积输出与全局描述 x_g 逐元素相乘[41];而 3×3 分支则采用核大小为 3×3 一维卷积来捕获更大感受野下的局部特征,同样与 x_g 进行逐元素相乘,以实现全局与局部信息的融合。

接着进行与 CA 类似的处理,我们将两个编码特征在时间维度上连接起来,使其共享相同的一维卷积,而不会在一维卷积分支中进行降维。在将一维卷积的输出分解为两个向量后,采用两个非线性 Sigmoid 函数来拟合线性卷积的二维二项式分布[39]。为了在一维卷积分支中的两条平行路线之间实现不同的跨通道交互功能,我们通过简单的乘法将每组内的两个通道注意力图聚合在一起。对于另一边,采用 3×3 的一维卷积分支,通过捕获局部跨信道交互来扩大特征空间。这样,EMA 不仅对通道间的信息进行编码以调整不同通道的重要性,而且将精确的时序结构信息保留到通道中。

最后,将 1×1 分支的两条路线与 3×3 分支输出相加,通过 Softmax 函数在时序维度上对融合后的特征进行归一化,生成注意力权重。该注意力权重随后与原始分组特征逐元素相乘,并最终将分组后的特征重构回原始尺寸[42]。这样,EMA 不仅对各通道间的信息进行编码以调整不同通道的重要性,而且将精确的时序结构信息有效地保留并增强到通道特征中。通过上述过程,我们对碳价的长距离依赖关系进行建模,并将精确的位置信息嵌入到 EMA 中。融合不同尺度的上下文信息使神经网络模型能够对碳价数据产生更好的关注。

3.3. ASPP (Atrous Spatial Pyramid Pooling)

ASPP (Atrous Spatial Pyramid Pooling) [43]结合了空洞卷积(Atrous Convolution)和空间金字塔池化(Spatial Pyramid Pooling) [44]技术,旨在通过多尺度特征处理来增强模型对复杂上下文信息的捕捉能力,广泛应用于图像分割等任务。在处理碳价预测任务时,我们把 2D ASPP 中的卷积操作调整为 1D 卷积,目的是让模型能够适应时间序列数据,且为了提高 ASPP (Atrous Spatial Pyramid Pooling)模型的灵活性,使用 CondConv 替换掉原本的普通卷积,在保证多尺度特征提取的同时,增强模型对碳价数据中关键时间模式的敏感性和泛化能力,这使得模型在处理时间序列预测任务时,表现出更强的能力,更适合处理碳价预测任务。

碳价格预测任务容易受短期波动和长期趋势的共同影响,单一时间窗口的模型可能忽略跨尺度的依赖关系。常规的 MaxPool 操作虽然能够提取碳价序列中的显著波动特征,但仅保留局部窗口内的最大值

会导致其他关键信息的丢失,影响预测模型对价格的预测精度和长期趋势的建模能力。而且固定单一的卷积结构难以动态捕捉碳价序列中的多尺度细节,这会限制模型对碳价短期突变与长期结构性变化的联合表征能力。因此,我们尝试通过动态卷积与自适应池化操作,增强模型对碳价序列中局部扰动与全局趋势的协同建模能力。具体逻辑如下:

首先,输入的碳价格时间序列数据的形状为: $[B, C, L]$, B 为批量大小, C 为通道数, L 为时间步长,对输入数据恩地每个通道进行全局平均池化处理,将整个时间维度的信息浓缩为一个向量。通过这一步我们对整个序列的统计信息进行一个初步的提取,方便后续捕捉长期趋势,对于每 b 个样本和每个通道 c 来说,全局平均池化可用下面的公式表示:

$$y_{b,c} = \frac{1}{L} \sum_{i=1}^L x_{b,c,i}$$

在经过全局平均池化的初步处理后,我们将其分别输入至 5 条支路进行操作,为了提高模型的灵活性,在第一条支路我们选择使用动态卷积来代替常规大小的卷积,可以根据每个样本的全局特征动态生成卷积核,从而使得提取出的全局特征更加针对当前样本的特性,减少了固定卷积核可能带来的泛化不足。动态卷积预设了多个“专家”卷积核,每个专家都有一组固定参数,能够学习不同的特征模式。对于全局平均池化生成的向量,经过扁平化、dropout 和全连接层生成路由权重:

$$r_c = \sigma(W_{fc} \cdot \text{dropout}(\text{flatten}(x)))$$

其中, $r = [r_1, r_2, \dots, r_E]$ 为长度为 E (专家数量) 的向量。 σ 为激活函数。

接着对动态卷积核进行构造,给定每个专家的卷积核 W_e , 动态卷积核 K 是各专家卷积核的加权和:

$$K = \sum_{e=1}^E r_e W_e$$

最后使用生成的动态卷积核对输入进行卷积操作,得到输出特征:

$$y = \text{Conv}(x, K)$$

我们将这四条支路搭建为构建我们的池化金字塔。对于非动态卷积的支路,我们给定的该普通卷积一个膨胀因子 D , 叠加相应的空洞卷积层,提取不同尺度下的特征,空洞卷积可表示成如下公式:

$$y[i] = \sum_{k=0}^{K-1} x[i + D \cdot k] \cdot \omega[k]$$

其中, $y[i]$ 表示输出在位置 i 的值, x 是输入, $\omega[k]$ 是卷积核的第 k 个权重, K 是卷积核的大小。

最后对各支路的输出叠加后,并通过输出层的 CondConv $\rightarrow$ BN $\rightarrow$ ReLU $\rightarrow$ Dropout 降维至给定通道数。通过使用 CondConv 替换掉原本的普通卷积,在保证多尺度特征提取的同时,增强模型对碳价数据中关键时间模式的敏感性和泛化能力,得到最终结果。

通过上述操作,空洞空间卷积金字塔池化能够有效地捕捉输入序列中的全局趋势与局部细节。它首先通过全局池化提取全局统计信息,再利用条件卷积生成针对每个样本定制化的卷积核;同时,多尺度空洞卷积以不同膨胀率捕捉多尺度特征,最后自适应融合各分支信息,从而兼顾长期趋势与短期波动,显著提升碳价预测任务的鲁棒性与精度。

3.4. 模型的构建

关于碳价预测模型,本文构建了一个复合模型用于碳价预测。以优化器为 Sophia 的 LSTM 模型,引入了 Efficient Multi-Scale Attention (EMA) 和 Atrous Spatial Pyramid Pooling (ASPP) 模块,最后结合集成学

习 XGBoost。LSTM 用于捕捉时间序列数据中的长期依赖；EMA 增强了模型在多尺度上的特征关注；ASPP 能够有效地捕捉输入序列中的全局趋势与局部细节，扩展了感受野；而 XGBoost 则用于微调和提升预测精度。利用全国碳市场收盘价数据，搭建组合预测模型来预测碳价未来值，具体预测流程见图 2。

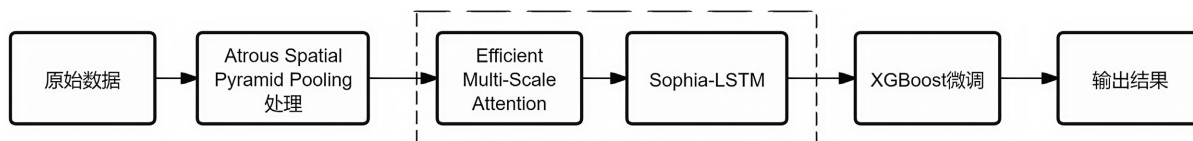


Figure 2. Combined model prediction flowchart

图 2. 组合模型预测流程图

原始数据序列共有 9 个维度，分别是全国碳市场开盘价、收盘价、最高价、最低价、成交量、沪深 300 指数、美元汇率、EUA 期货价格和中证新能源指数。实验采用了 Pytorch 框架，并部署了相关的预测模型，其中优化算法采用的 Sophia 算法，Sophia 算法比传统的优化算法具有更好的训练效果。数据按照 7:3 的比例划分成训练集和测试集。

具体训练过程中，数据序列经过 ASPP 模块预处理后，输入到基于序列到序列的基于注意力机制 EMA 以 Sophia 为优化器的 LSTM (EMA-Sophia-LSTM) 模型，其中基于注意力机制 EMA 是编码器，基于注意力机制 EMA 的并行子结构有助于网络出现过多的顺序处理和过大的深度，可以捕获 LSTM 模型可能无法捕获的特征，而 LSTM 模型是解码器，可以更好地描述时间序列特性。数据序列首先会输入基于注意力机制 EMA 模型，学习当前局部序列特征的重要程度，然后输入 LSTM 模型，在 LSTM 模型中，采用 10 天滞后的时间窗口进行滚动预测，由 LSTM 模型对长期序列进行学习，更全面地刻画时序数据。

而后进入了 XGBoost 微调，解码后，也即得到了经过机器学习模型预测的碳价序列，此时再将数据序列输入 XGBoost 模型。XGBoost 模型集成了多个决策树模型，通过迭代树模型来改进预测能力。将机器学习模型预测后得到的碳价序列输入到 XGBoost 模型中，可以视为进行二次预测或微调，这样做结合了两模型的优点，可以捕捉到更多信息和模式，最终得到预测结果。

组合模型以碳价收盘价作为输出结果，对比实际全国碳市场收盘价实际价格，以均方误差(MSE)、均方根误差(RMSE)、平均绝对误差(MAE)和 R^2 值作为量化模型效果的指标。MSE、RMSE、MAE 值越小， R^2 值越大，表示模型的预测效果更好。

4. 实验与分析

4.1. 实验过程

为了验证上述复合模型(ASPP-EMA-LSTM-XGBoost)的准确性，本文分别对比了 LSTM 模型、ASPP-LSTM 模型、EMA-LSTM 模型、Sophia-LSTM 模型以及 LSTM-XGBoost 的表现。通过这些对比，可评估每个组件对模型性能的影响，并验证最终复合模型的优越性。

4.2. 实验结果与分析

为了验证组合模型有效性，本文还选取了组合模型的单一模型 LSTM 及与 LSTM 单一组合模块模型，即 LSTM 模型、ASPP-LSTM 模型、EMA-LSTM 模型、Sophia-LSTM 模型以及 LSTM-XGBoost 作为对比模型，以模型的 MSE 值、RMSE 值、MAE 值和 R^2 值作为评判模型优劣的标准。预测模型的结果见下图 3~图 8。

模型预测的性能指标见表 3。

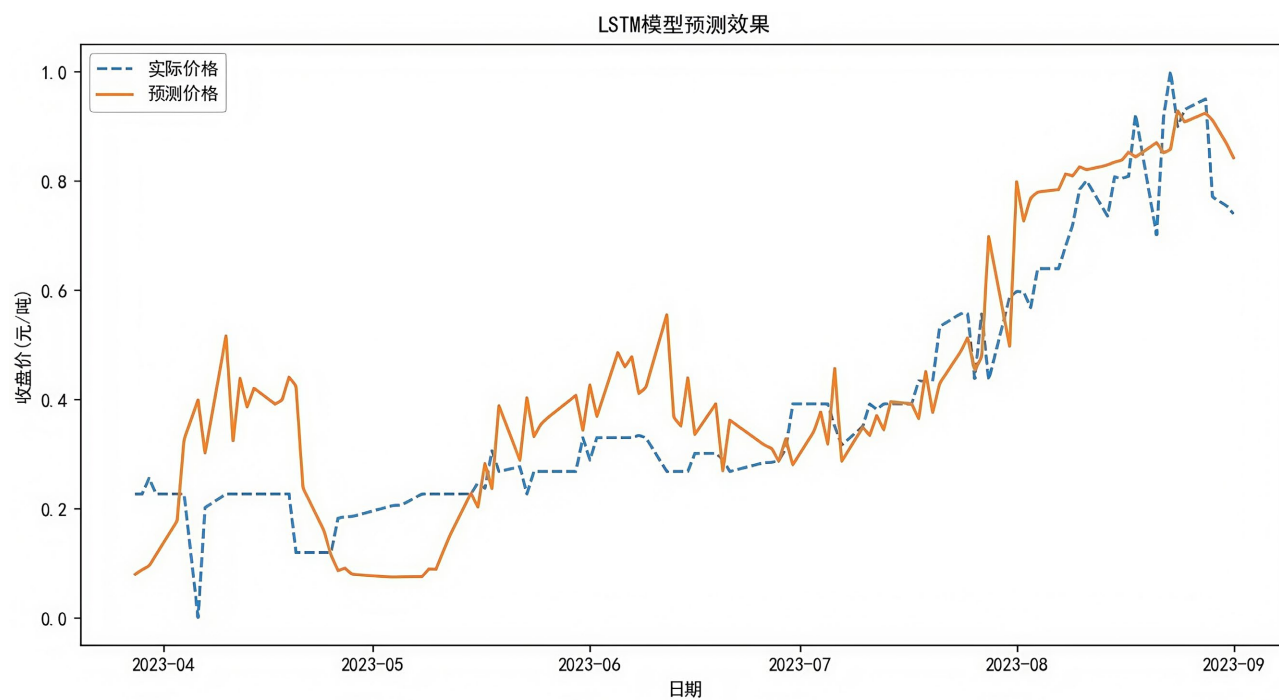


Figure 3. LSTM model prediction results

图 3. LSTM 模型预测结果

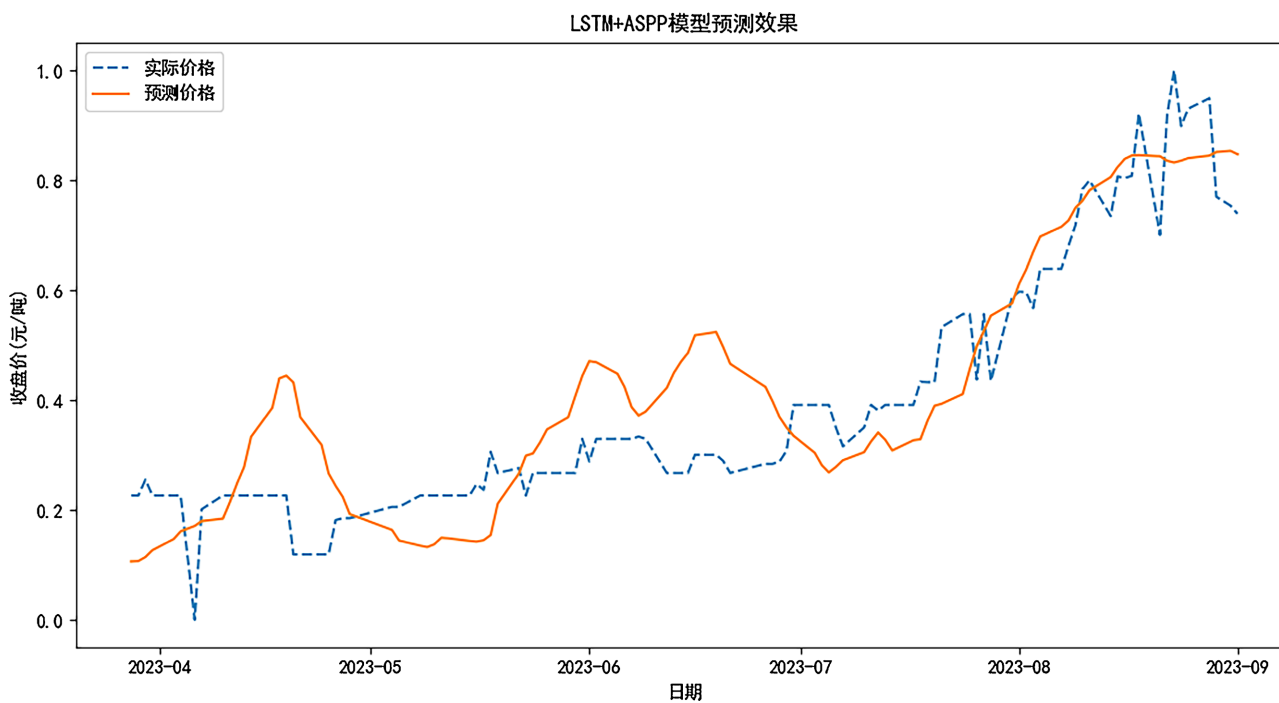


Figure 4. ASPP-LSTM model prediction results

图 4. ASPP-LSTM 模型预测结果

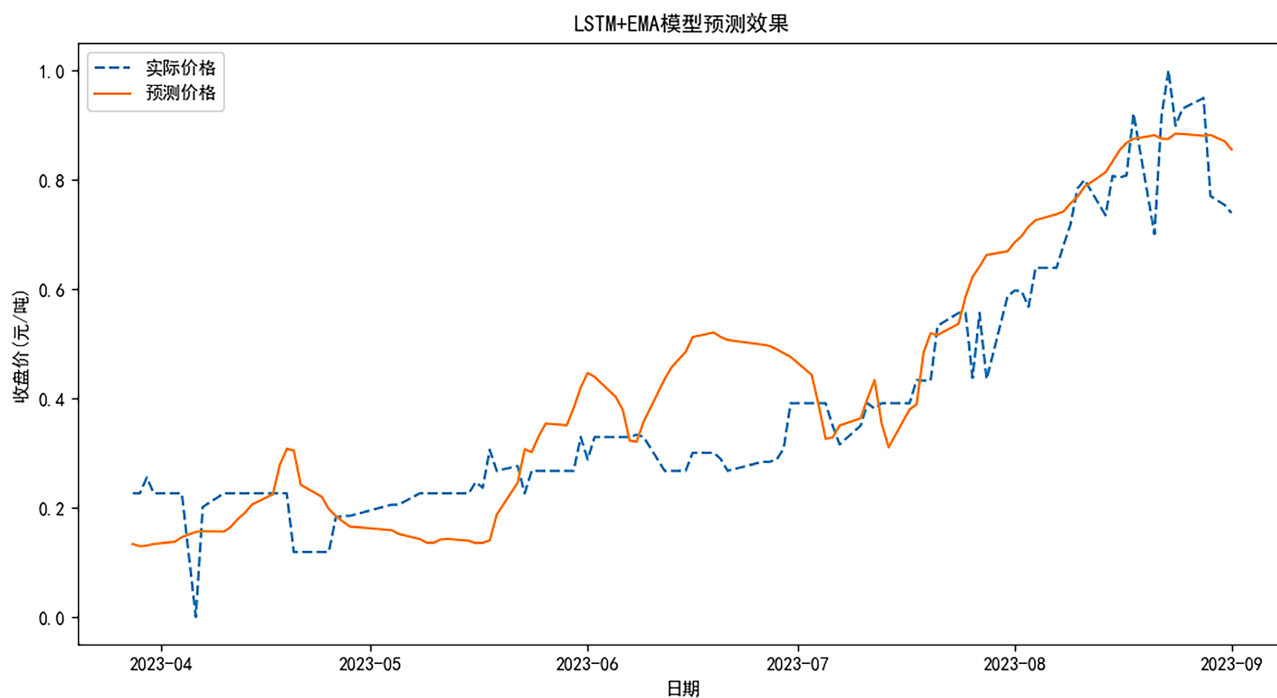


Figure 5. EMA-LSTM model prediction results

图 5. EMA-LSTM 模型预测结果

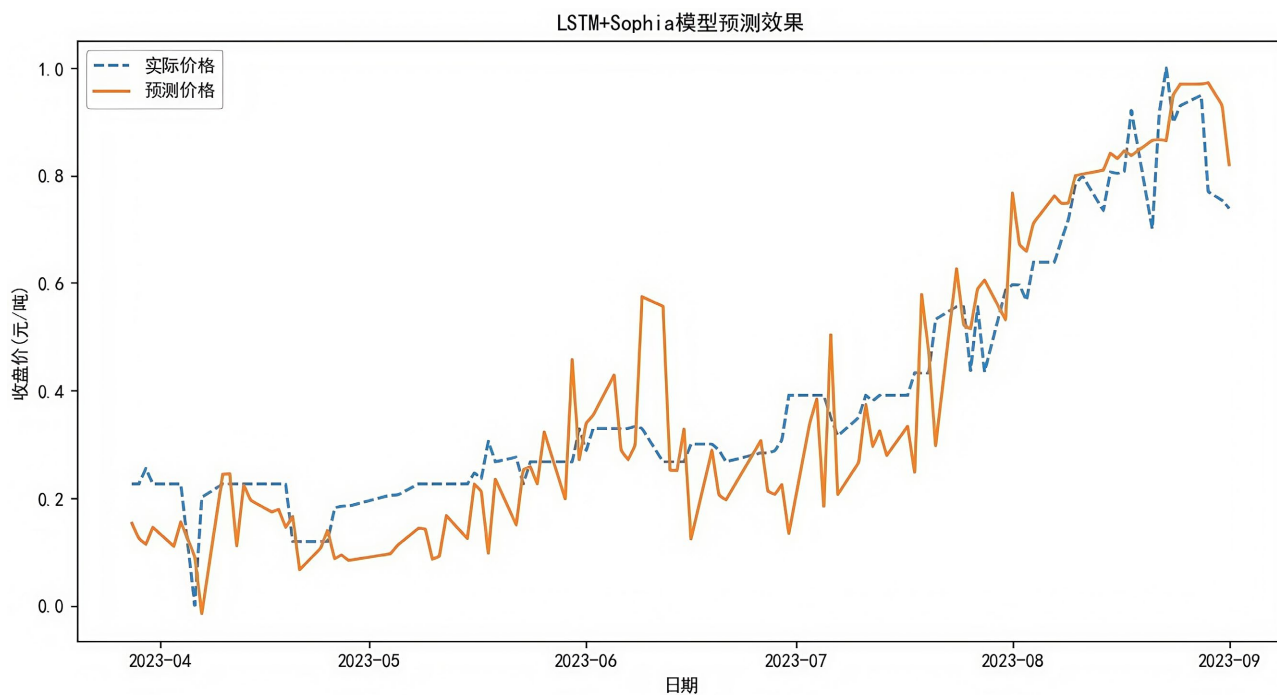


Figure 6. Sophia-LSTM model prediction results

图 6. Sophia-LSTM 模型预测结果

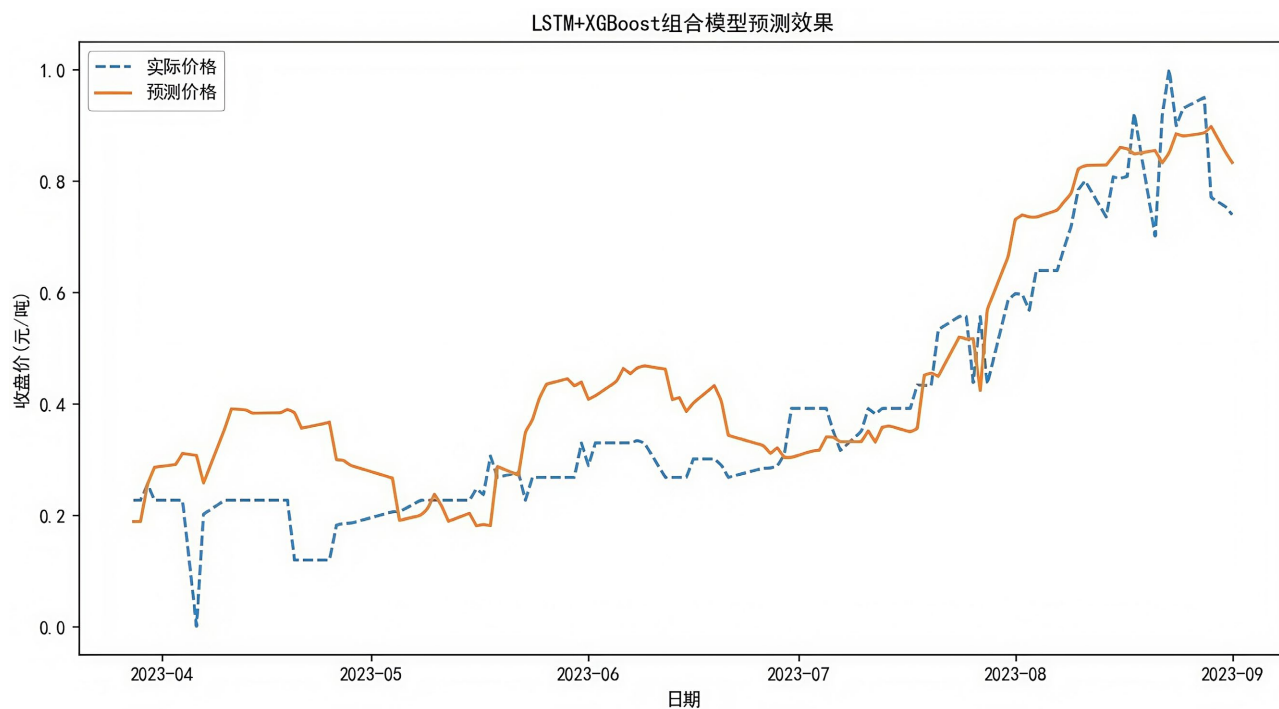


Figure 7. LSTM-XGBoost model prediction results

图 7. LSTM-XGBoost 模型预测结果

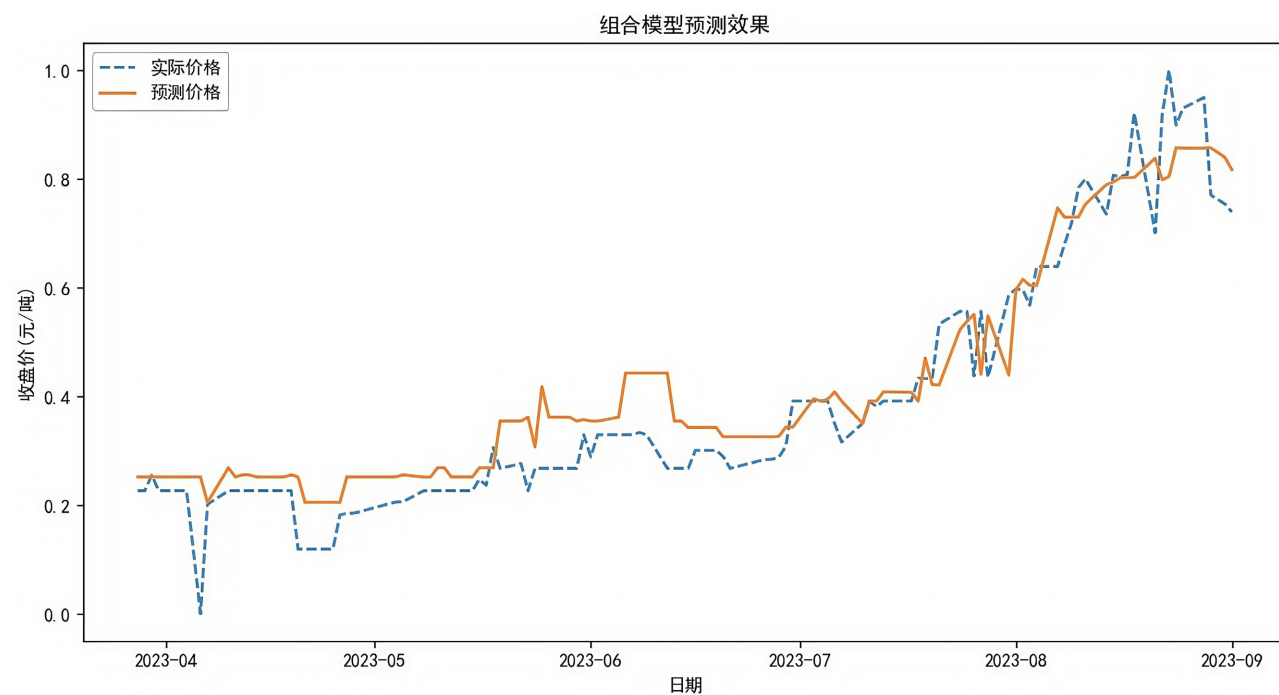


Figure 8. Combined model prediction results

图 8. 组合模型预测结果

对比模型预测性能指标可以发现：(1) 组合模型的 MSE、RMSE 和 MAE 值均小于其余五个模型，这说明组合模型相较单一模型可以更好地减低模型的预测误差；(2) 组合模型的 R^2 值较 LSTM 模型提高了

29.74%，较 EMA-LSTM 模型提高了 17.31%，较 ASPP-LSTM 模型提高了 21.98%，较 Sophia-LSTM 模型提高了 16.14%，较 LSTM-XGBoost 模型提高了 20.88%。这说明组合模型的预测准确率要优于单一模型，由此表明了组合模型的有效性，以及组合模型较单一模型的预测优势。

Table 3. Model performance evaluation metrics
表 3. 模型性能评价指标

模型	MSE	RMSE	MAE	R ²
LSTM	0.0150	0.1225	0.0989	0.7032
LSTM + EMA	0.0112	0.1058	0.0866	0.7777
LSTM + ASPP	0.0127	0.1127	0.0952	0.7479
LSTM + Sophia	0.0108	0.1039	0.0838	0.7855
LSTM + XGBoost	0.0124	0.1114	0.0923	0.7547
组合模型	0.0044	0.0663	0.0476	0.9123

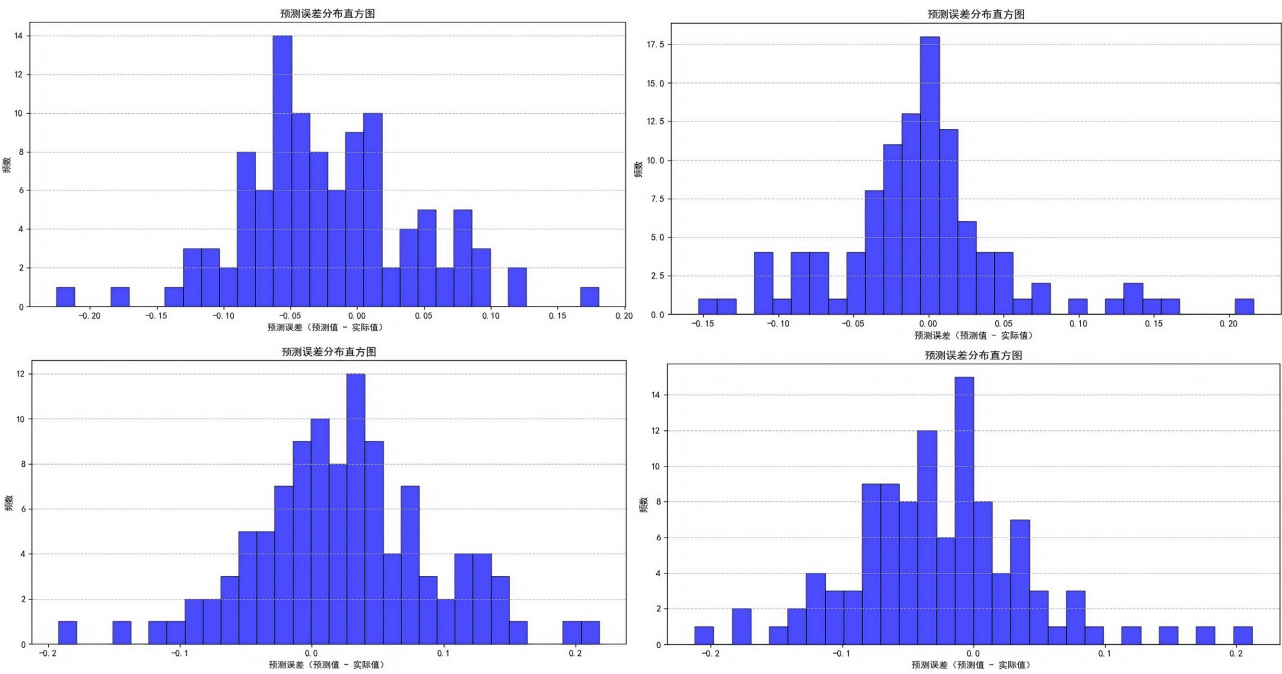


Figure 9. Error analysis of combinatorial models
图 9. 组合模型的误差分析

组合模型的误差分析如图 9。图 9 为组合模型的误差分析。分别展示了模型在 4 次运行中的表现。尽管每次运行时误差分布有所波动，但总体来看误差值大多集中在零附近，且分布形态保持一致。这种波动性是许多机器学习模型包含随机成分时的常见现象。因此，误差的波动并不意味着模型不可靠，反而体现了其鲁棒性。这表明，尽管存在一定的随机性，模型的预测能力在不同实验之间是稳定的。

5. 结论

本文选取全国碳市场的碳价作为研究对象，首先从理论分析碳价的影响因素，然后采用皮尔逊相关

系数法提取碳价关键影响要素, 对比组合模型及其单一模型 LSTM 及与 LSTM 单一组合模块模型, 即 LSTM 模型、ASPP-LSTM 模型、EMA-LSTM 模型、Sophia-LSTM 模型以及 LSTM-XGBoost 作为对比模型, 通过 MSE、RMSE、MAE 和 R^2 指标检验预测效果, 以及对组合模型的误差分析发现注意力机制组合模型具有更高的预测准确率, 表明了组合模型的有效性以及较单一预测模型的实用性。

参考文献

- [1] Bredin, D. and Muckley, C. (2011) An Emerging Equilibrium in the EU Emissions Trading Scheme. *Energy Economics*, **33**, 353-362. <https://doi.org/10.1016/j.eneco.2010.06.009>
- [2] Aatola, P., Ollikainen, M. and Toppinen, A. (2013) Price Determination in the EU ETS Market: Theory and Econometric Analysis with Market Fundamentals. *Energy Economics*, **36**, 380-395. <https://doi.org/10.1016/j.eneco.2012.09.009>
- [3] 张玲, 曹峰. 环境规制对碳交易价格波动的影响分析[J]. 价格月刊, 2023(1): 1-7.
- [4] 姜瑜, 吴哲宇. 我国碳排放权交易价格影响因素回归分析[J]. 环境与可持续发展, 2021, 46(1): 77-83.
- [5] 王小燕, 周思敏, 徐晓莉, 等. 基于图结构自适应 Lasso 的碳排放权价格影响因素分析[J]. 统计与信息论坛, 2022, 37(4): 73-83.
- [6] 吴慧娟, 张智光. 城市碳价的时空特征及其形成机理的理论模型——基于 8 个地区碳交易试点的价格数据[J]. 现代城市研究, 2021, 36(1): 19-24.
- [7] Zhang, J. and Xu, Y. (2020) Research on the Price Fluctuation and Risk Formation Mechanism of Carbon Emission Rights in China Based on a GARCH Model. *Sustainability*, **12**, Article 4249. <https://doi.org/10.3390/su12104249>
- [8] Jiang, L. and Wu, P. (2016) International Carbon Market Price Forecasting Research Based on ARIMA-RF Model. *Proceedings of the 2016 International Conference on Strategic Management*, Sichuan, November 2015, 1089-1092.
- [9] 王娜. 基于 Boosting-ARMA 的碳价预测[J]. 统计与信息论坛, 2017, 32(3): 28-34.
- [10] Zhu, B., Han, D., Wang, P., Wu, Z., Zhang, T. and Wei, Y. (2017) Forecasting Carbon Price Using Empirical Mode Decomposition and Evolutionary Least Squares Support Vector Regression. *Applied Energy*, **191**, 521-530. <https://doi.org/10.1016/j.apenergy.2017.01.076>
- [11] 金林, 马忠芸, 王红红. 基于灰色 BP 神经网络碳排放交易价格预测[J]. 河北环境工程学院学报, 2020, 30(1): 27-32, 41.
- [12] Li, H., Huang, X., Zhou, D., Cao, A., Su, M., Wang, Y., et al. (2022) Forecasting Carbon Price in China: A Multimodel Comparison. *International Journal of Environmental Research and Public Health*, **19**, Article 6217. <https://doi.org/10.3390/ijerph19106217>
- [13] 朱亮亮, 肖楚鹏, 余梦, 等. 基于 PCC-LSTM 模型的全国碳排放权交易价格预测研究[J]. 环境保护科学, 2023, 49(5): 55-62, 129.
- [14] 呼雪芳. 基于 CEEMDAN 分解的碳交易价格组合预测模型研究[D]: [硕士学位论文]. 广州: 华南理工大学, 2021.
- [15] 高长征, 李东伟, 王秀娜, 等. 利用智能机器学习方法对区域碳排放权交易价格预测研究——基于湖北碳市场数据的分析[J]. 价格理论与实践, 2022(4): 89-93, 205.
- [16] Wang, J., Zhu, H., Cheng, F., Zhou, C., Zhang, Y., Xu, H., et al. (2023) A Novel Wind Power Prediction Model Improved with Feature Enhancement and Autoregressive Error Compensation. *Journal of Cleaner Production*, **420**, Article 138386. <https://doi.org/10.1016/j.jclepro.2023.138386>
- [17] Du, X., Sun, Y., Zhang, X. and Zhang, R. (2023) Research on the Impact of Energy Price on Carbon Emission Intensity of China—An Empirical Study Based on LMDI Decomposition and Econometric Models. *Sustainability*, **15**, Article 8528. <https://doi.org/10.3390/su15118528>
- [18] Liu, H., Zhao, T., Wang, S. and Li, X. (2023) A Stock Rank Prediction Method Combining Industry Attributes and Price Data of Stocks. *Information Processing & Management*, **60**, Article 103358. <https://doi.org/10.1016/j.ipm.2023.103358>
- [19] 洪涓, 陈静. 我国碳交易市场价格影响因素分析[J]. 价格理论与实践, 2009(12): 65-66.
- [20] Ge, Q. (2025) Enhancing Stock Market Forecasting: A Hybrid Model for Accurate Prediction of S&P 500 and CSI 300 Future Prices. *Expert Systems with Applications*, **260**, Article 125380. <https://doi.org/10.1016/j.eswa.2024.125380>
- [21] Galati, G. and Ho, C. (2003) Macroeconomic News and the Euro/Dollar Exchange Rate. *Economic Notes*, **32**, 371-398. <https://doi.org/10.1111/1468-0300.00118>
- [22] Arlinghaus, J. (2015) Impacts of Carbon Prices on Indicators of Competitiveness: A Review of Empirical Findings. OECD Environment Working Papers 87, OECD Publishing.

- [23] Sun, C., Ding, D., Fang, X., Zhang, H. and Li, J. (2019) How Do Fossil Energy Prices Affect the Stock Prices of New Energy Companies? Evidence from Divisia Energy Price Index in China's Market. *Energy*, **169**, 637-645. <https://doi.org/10.1016/j.energy.2018.12.032>
- [24] Hambling, T., Weinstein, P. and Slaney, D. (2011) A Review of Frameworks for Developing Environmental Health Indicators for Climate Change and Health. *International Journal of Environmental Research and Public Health*, **8**, 2854-2875. <https://doi.org/10.3390/ijerph8072854>
- [25] Pratama, I., Permanasari, A.E., Ardiyanto, I. and Indrayani, R. (2016) A Review of Missing Values Handling Methods on Time-Series Data. 2016 *International Conference on Information Technology Systems and Innovation (ICITSI)*, Bandung, 24-27 October 2016, 1-6. <https://doi.org/10.1109/icitsi.2016.7858189>
- [26] Ji, C., Hu, Y. and Tang, B. (2018) Research on Carbon Market Price Mechanism and Influencing Factors: A Literature Review. *Natural Hazards*, **92**, 761-782. <https://doi.org/10.1007/s11069-018-3223-1>
- [27] Abounaima, M.C., Mazouri, F.Z.E., Lamrini, L., Nfissi, N., Makhfi, N.E. and Ouzarf, M. (2020) The Pearson Correlation Coefficient Applied to Compare Multi-Criteria Methods: Case the Ranking Problematic. 2020 *1st International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET)*, Meknes, 16-19 April 2020, 1-6. <https://doi.org/10.1109/iraset48871.2020.9092242>
- [28] Liu, H., Li, Z., Hall, D., *et al.* (2023) Sophia: A Scalable Stochastic Second-Order Optimizer for Language Model Pre-training. arXiv:2305.14342.
- [29] Hochreiter, S. and Schmidhuber, J. (1997) Long Short-Term Memory. *Neural Computation*, **9**, 1735-1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- [30] Kingma, D.P. and Ba, J. (2014) Adam: A Method for Stochastic Optimization. arXiv:1412.6980.
- [31] Galántai, A. (2000) The Theory of Newton's Method. *Journal of Computational and Applied Mathematics*, **124**, 25-44. [https://doi.org/10.1016/S0377-0427\(00\)00435-0](https://doi.org/10.1016/S0377-0427(00)00435-0)
- [32] Klinker, F. (2011) Exponential Moving Average versus Moving Exponential Average. *Mathematische Semesterberichte*, **58**, 97-107. <https://doi.org/10.1007/s00591-010-0080-8>
- [33] Wang, Y., Qin, L., Wang, Q., Chen, Y., Yang, Q., Xing, L., *et al.* (2023) A Novel Deep Learning Carbon Price Short-Term Prediction Model with Dual-Stage Attention Mechanism. *Applied Energy*, **347**, Article 121380. <https://doi.org/10.1016/j.apenergy.2023.121380>
- [34] Lin, X., Zhu, X., Feng, M., Han, Y. and Geng, Z. (2021) Economy and Carbon Emissions Optimization of Different Countries or Areas in the World Using an Improved Attention Mechanism Based Long Short Term Memory Neural Network. *Science of The Total Environment*, **792**, Article 148444. <https://doi.org/10.1016/j.scitotenv.2021.148444>
- [35] Niu, Z., Zhong, G. and Yu, H. (2021) A Review on the Attention Mechanism of Deep Learning. *Neurocomputing*, **452**, 48-62. <https://doi.org/10.1016/j.neucom.2021.03.091>
- [36] Johnston, W.A. and Dark, V.J. (1986) Selective Attention. *Annual Review of Psychology*, **37**, 43-75. <https://doi.org/10.1146/annurev.ps.37.020186.000355>
- [37] Woo, S., Park, J., Lee, J. and Kweon, I.S. (2018) CBAM: Convolutional Block Attention Module. In: Ferrari, V., Hebert, M., Sminchisescu, C. and Weiss, Y., Eds., *Lecture Notes in Computer Science*, Springer International Publishing, 3-19. https://doi.org/10.1007/978-3-030-01234-2_1
- [38] Hou, Q., Zhou, D. and Feng, J. (2021) Coordinate Attention for Efficient Mobile Network Design. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, 20-25 June 2021, 13708-13717. <https://doi.org/10.1109/cvpr46437.2021.01350>
- [39] Ouyang, D., He, S., Zhang, G., Luo, M., Guo, H., Zhan, J., *et al.* (2023) Efficient Multi-Scale Attention Module with Cross-Spatial Learning. *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Rhodes Island, 4-10 June 2023, 1-5. <https://doi.org/10.1109/icassp49357.2023.10096516>
- [40] Li, X., Hu, X. and Yang, J. (2019) Spatial Group-Wise Enhance: Improving Semantic Feature Learning in Convolutional Networks. arXiv:1905.09646.
- [41] Liu, T., Luo, R., Xu, L., Feng, D., Cao, L., Liu, S., *et al.* (2022) Spatial Channel Attention for Deep Convolutional Neural Networks. *Mathematics*, **10**, Article 1750. <https://doi.org/10.3390/math10101750>
- [42] Chen, Y., Kalantidis, Y., Li, J., *et al.* (2018) A²-Nets: Double Attention Networks. *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, Montréal, 3-8 December 2018, 350-359.
- [43] Chen, L., Papandreou, G., Kokkinos, I., Murphy, K. and Yuille, A.L. (2018) Deeplab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **40**, 834-848. <https://doi.org/10.1109/tpami.2017.2699184>
- [44] He, K., Zhang, X., Ren, S. and Sun, J. (2015) Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **37**, 1904-1916. <https://doi.org/10.1109/tpami.2015.2389824>