

基于趋势性时间序列的全国碳排放量预测研究

钟进, 李宗航

贵州大学数学与统计学院, 贵州 贵阳

收稿日期: 2023年7月4日; 录用日期: 2023年8月10日; 发布日期: 2023年8月17日

摘要

根据最新数据和趋势分析, 中国碳排放量一直处于持续增长的形势, 尽管中国政府已经采取推广清洁能源、加强能源效率、推动碳交易市场建设等一系列措施以达到节能减排目的, 但中国碳排放量仍然面临诸多问题。因此本文基于2019年1月至2023年3月的1551条日频数据, 利用SARIMA、LSTM以及SVR等模型综合考虑时间序列的趋势性, 对比分析了不同模型对于碳排放数据的预测效果, 结果表明SARIMA模型对于碳排放数据的预测效果优于其余两个模型。

关键词

碳排放, 时间序列, LSTM模型, SARIMA模型, SVR模型

Research on National Carbon Emission Forecasting Based on Trend Time Series

Jin Zhong, Zonghang Li

School of Mathematics and Statistics, Guizhou University, Guiyang Guizhou

Received: Jul. 4th, 2023; accepted: Aug. 10th, 2023; published: Aug. 17th, 2023

Abstract

According to the latest data and trend analysis, China's carbon emissions have been in a situation of continuous growth, and although the Chinese government has taken a series of measures to promote clean energy, strengthen energy efficiency, and promote the construction of carbon trading market to achieve energy conservation and emission reduction, China's carbon emissions still face many problems. Therefore, in this paper, based on 1551 daily frequency data from January 2019 to March 2023, we compare and analyze the prediction effect of different models for carbon emission data using SARIMA, LSTM and SVR models considering the trend of time series,

and the results show that SARIMA model has better prediction effect than the remaining two models for carbon emission data.

Keywords

Carbon Emissions, Time Series, LSTM Model, SARIMA Model, SVR Model

Copyright © 2023 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

随着全球经济的快速发展和人口规模的增加,碳排放成为全球气候变化的主要原因之一。碳排放的增加不仅导致全球气温升高和极端天气的增多,还对生态环境和人类健康造成了重大威胁。作为全球最大的碳排放国之一,中国的碳排放量一直处于全球领先地位,根据国际能源署的数据,2019年中国二氧化碳排放量达到了10.17亿吨,占全球总排放量的28%。为了应对气候变化挑战,中国政府采取了一系列措施来减少碳排放量。2020年9月第七十五届联合国大会上国家主席首次明确提出碳达峰和碳中和的目标,承诺力争于2030年前达到峰值,2060年前实现碳中和的宏远目标。2021年10月国务院发布《2030年前碳达峰行动方案》,方案中提出将碳达峰贯穿于经济社会发展全过程和各方面,重点实施能源绿色低碳转型行动、城乡建设碳达峰行动、各地区梯次有序碳达峰行动等十大行动。尽管中国政府采取多项并行的举措以实现碳达峰目标,然而中国的碳排放量仍然面临许多挑战和压力,碳排放量增加的趋势仍然存在,中国在碳排放领域存在的一些不足,比如,高碳能源依赖——中国仍然依赖煤炭等高碳能源,这导致了较高的温室气体排放量;工业排放控制不足——工业部门是中国碳排放的主要来源之一,尤其是钢铁、水泥等重工业行业;建筑和交通行业碳排放问题——建筑和交通行业在中国的碳排放量也相当大,特别是在城市化进程中,与城市发展相关的建筑和交通需求也带来了碳排放的增加了的压力;农业和森林保护——农业和森林保护领域也存在一些挑战,例如农用地的大规模开垦导致的碳排放和森林砍伐等;监测和数据不足——在中国,碳排放的准确监测和数据收集仍然存在一定的不足。要解决这些不足,中国需要进一步加强政策制定和执行,加大技术创新和投入,加强国际合作,推动低碳发展和绿色转型。鉴于此,为了更好地制定和实施碳减排政策,需要建立完善的碳排放数据监测体系,提高数据的准确性和及时性。本文基于2019年1月~2023年3月全国碳排放日频数据分析对比了不同模型的预测效果,对于未来中国碳排放量的预测具有重要的现实意义和科学价值,促进中国的低碳可持续发展。

2. 文献综述

在全球碳排放数据不断增加的环境下,国内外学术界针对碳排放问题进行了研究,主要集中在碳排放影响因素以及各行业碳排放量预测两个方面。

在研究碳排放影响因素方面,主要研究集中于指数分解法、改进的STIRPAT模型、回归模型等影响因素分析方法。Kaya恒等式和LMDI分解法作为指数分解法的代表被广大学者用于碳排放影响因素分析,冯相昭 and 邹骥(2008) [1]对Kaya恒等式进行改进后对中国的CO₂排放量进行了无残差分解,得出经济驱动和人口增长是碳排放增加的主要驱动因素。袁路和潘家华(2013) [2]对Kaya恒等式对碳排放影响因素对发展现状及其存在对局限性进行了详细阐述。Pan Jiang & Xiu Juan等(2023) [3]基于LMDI模型对中国八

个经济区域碳排放对时空差异及影响因素进行了研究。张矢宇、杨杰等(2023) [4]通过 LDMI 分解法得出影响我国交通碳排放的主要因素为交通运输结构、单位运输能耗水平、城镇居民人均 GDP 等。STRIPAT 作为从 IPAT 模型中改进得到的模型在近年来多用于碳排放分析, 胡广阔、李春梅和惠树鹏(2016) [5]通过贝叶斯法对于 STRIPAT 模型的解释变量重要性进行筛选, 验证了我国碳排放强度与选取解释变量的关联影响。刘茂辉、翟华欣等(2023) [6]通过对比 LMDI 模型和 STRIPAT 模型分析了天津市的碳排放量, 并从基准、低碳和超低碳三种情景预测了碳达峰和碳中和的情况。除此之外, 杨青、彭若慧等(2023) [7]基于地理加权回归分析了我国不同省份碳排放量情况, 揭示了碳排放量与区域社会经济发展的关系。

在各行业碳排放量的预测上, 肇晓楠、谢新连等(2023) [8]在利用灰色关联度获取主要影响因素的基础上通过滑动窗口动态输入 LSTM 网络对铁路运输碳排放量进行了预测。池小波、续泽晋等(2023) [9]提出一种分量增广输入的 WPD-ISSA-CA-CNN 碳排放预测模型, 利用该模型以山西某发电厂的历史数据为样本对电厂碳排放进行了预测, 得出该模型在火力发电碳排放中预测效果交好的结论。Aryai Vahid & Goldsworthy Mark (2023) [10]基于粒子群优化-极端随机树(PSO-ERT)回归器模型对澳大利亚国家电力市场的碳排放强度进行了预测。Zhao Jinjie & Kou Lei 等人(2022) [11]采用二次分配程序回归分析, 从区域差异的角度分析了影响黄河流域碳排放的因素, 并通过麻雀搜索算法优化的长短时记忆网络(SSA-LSTM)对黄河流域的碳排放量进行了预测。J B Yuan & Y S Liu (2019) [12]将民用建筑分为城市住宅建筑、农村住宅建筑、公共建筑和北部城市供暖, 根据改进的灰色预测方法构建碳排放模型, 预测了 2017 年至 2025 年全国民用建筑总碳排放量。

基于以上对于现有碳排放相关文献的分析, 国内外对于碳排放的研究已趋于成熟, 并且主要集中于碳排放影响因素分析和不同行业的碳排放量预测等方面。本文主要集中于对碳排放量时间序列的分析, 基于滑动窗口的思想利用 SARIMA、LSTM、SVR 等模型对 2019~2023 年全国二氧化碳日排放量进行研究, 最后通过评估指标分析不同模型对碳排放量预测的效果。

3. 碳排放预测模型理论

3.1. 季节性差分自回归移动平均模型 SARIMA

SARIMA 模型(Seasonal Auto-regressive Integrated Moving Average)是将季节差分与差分自回归移动平均模型(ARIMA)模型相结合的用于处理具有周期性特征的时间序列数据的模型(1997) [13]。应用于包含趋势和季节性的单变量数据, SARIMA 由趋势和季节要素组成的序列构成。

自回归(AR)是描述当前值与历史值之间的关系, 利用变量自身的历史时间数据对其进行预测, p 阶自回归过程公式可表示为:

$$y_t = \alpha + \sum_{i=1}^p \gamma_i y_{t-i} + \varepsilon_t$$

其中, y_t 是当前值, α 是常数项, p 是阶数, γ_i 是自相关系数, ε_t 是误差项。

数据差分(I)是指将非平稳数据转换为平稳的过程, 去除其非恒定的趋势。当时间序列 Y_t 不平稳时需要按照 d 阶差分公式进行差分:

$$u_t = \Delta^d y_t = (1 - B)^d y_t$$

其中, u 为差分算子, d 为差分阶数。

移动平均模型(MA)主要考虑的是自回归模型中的误差项的累加, 以有效消除预测中的随机波动。 q 阶自回归过程公式可表示为:

$$y_t = \alpha + \varepsilon_t + \sum_{i=1}^q \Theta_i \varepsilon_{t-i}$$

其中, ε_t 是 t 时刻的误差项。

ARIMA(p, d, q)模型通常是自回归(AR)、差分(I)和移动平均模型(MA)三部分的整合, 是统计模型中可用于非平稳时间序列预测的线性模型(2019) [14]。当原时间序列进行 d 阶差分转化为平稳序列后, 可以通过自相关系数(ACF)和偏自相关函数(PACF)或者计算赤池信息准则(AIC)和贝叶斯信息(BIC)的值确定具体的 p 和 q 阶数的取值, 如表 1 所示。

Table 1. ARIMA (p, d, q) order determination table

表 1. ARIMA(p, d, q)阶数确定表

模型	ACF	PACF
AR(p)	衰减趋于零 (几何型或振荡型)	p 阶后截尾
MA(q)	q 阶后截尾	衰减趋于零 (几何型或振荡型)
ARMA(p, q)	q 阶后衰减趋于零 (几何型或振荡型)	p 阶后衰减趋于零 (几何型或振荡型)

$$AIC = 2k - 2\ln(L)$$

$$BIC = k \ln(n) - 2\ln(L)$$

其中, k 为模型参数个数, n 为样本数量, L 为似然函数; 在保证模型精度的情况下尽量使得 k 值越小越好。

尽管 ARIMA 模型能够很好的处理非平稳时间序列并对其进行拟合, 但当原时间序列 Y_t 带有趋势性和季节性时, 仅仅利用 ARIMA 模型进行拟合会出现较大误差。SARIMA 模型利用季节差分对 ARIMA 模型进行优化能很好地拟合带有季节性的时间序列数据。SARIMA 模型参数可以分为非季节参数(p, d, q)和季节参数(P, D, Q, s)两大类(2016) [15]:

$$\text{SARIMA}(\underbrace{p, d, q}_{\text{non-seasonal}})(\underbrace{P, D, Q, s}_{\text{seasonal}})$$

其中, p 表示非季节自回归阶数, d 表示一步差分次数, q 表示非季节移动平均的阶数。季节参数中 P 表示季节自回归阶数, D 表示季节差分次数, Q 表示季节移动平均阶数, s 表示季节长度, 或者称为周期大小。

SARIMA (Seasonal Autoregressive Integrated Moving Average)模型是一种用于时间序列分析和预测的统计模型。SARIMA 模型的主要用途包括以下几个方面: 时间序列分析、时间序列预测、趋势和季节性调整、销售预测和需求预测、经济预测等, 需要注意的是, 它是 ARIMA 模型的扩展, SARIMA 模型的应用需要根据具体问题和数据的特点进行调整和定制, 模型的选择和参数设置也需要根据实际情况进行优化。

3.2. 长短时期记忆模型 LSTM

传统的循环神经网络(RNN)是一类以序列(sequence)数据为输入, 在序列的演进方向进行递归(recursion)且所有节点(循环单元)按链式连接的递归神经网络。利用 RNN 模型进行时间序列预测时无法捕

获序列的长时间依赖关系, 为解决这一问题, Hochreiter & Schmidhuber [16]于 1997 年提出了一种 RNN 的特殊形式—长短期记忆网络(LSTM), 如图 1 所示这一模型主要由遗忘门 f_t 、输入门 i_t 、细胞状态 \tilde{C}_t 和输出门 h_t 组成。

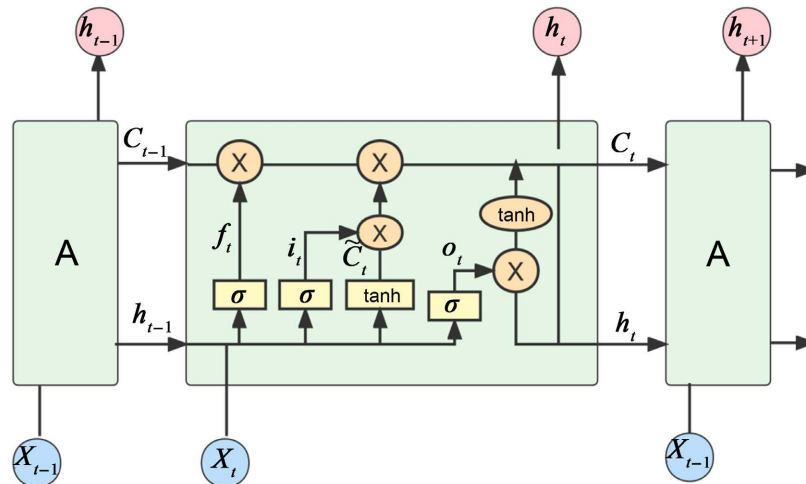


Figure 1. Schematic diagram of LSTM neural network
图 1. LSTM 神经网络结构示意图

LSTM 神经网络通常由三步骤实现:

第一步: 决定从单元状态中舍弃什么信息, 由激活函数为 sigmoid 的遗忘门 f_t 决定:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

σ 为激活函数 sigmoid, x_t 是当前输入向量, h_t 是当前隐含层向量, b_f 、 W_f 分别是偏置和输入权重。

第二步: 决定将要在单元状态(cell)中存储哪些新的信息, 其中激活函数为 sigmoid 的输入门 i_t 决定将更新哪些值, 激活函数为 tanh 的层 \tilde{C}_t 将创建一个新的候选值向量, 添加到单元状态:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

第三步: 输出门 h_t 将决定 LSTM 模型输出的内容:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$

$$h_t = o_t * \tanh(C_t)$$

经过以上步骤构建的 LSTM 模型能够很好的序列数据, 能够有效地处理长期依赖关系, 进而提高模型预测精度。

LSTM 是一种循环神经网络的变型, 它在处理序列数据时能够更好地捕捉长期依赖关系。LSTM 模型的应用非常广泛, 主要用途包括以下几个方面: 自然语言处理、语音识别、图像描述生成、时间序列预测、推荐系统等方面, 并且它能够有效处理时序数据的长期依赖关系, 提升模型的准确性和性能。

3.3. 支持向量回归 SVR

支持向量回归(SVR)是一种基于支持向量机(support vector machine, SVM)的回归分析方法, 用于预测

连续变量或实数值(2003) [17]。其主要思想是寻找训练数据中的支持向量, 来构建一个能够最好地拟合训练数据的超平面。与传统的回归方法不同, 支持向量回归使用核函数将数据映射到高维空间中, 从而使非线性关系也能够被拟合。对于回归函数:

$$f(x) = w\phi(x) + b$$

其中 w 为权值向量, b 为偏置项, SVR 问题在引入损失函数的情况下可形式化为:

$$\begin{aligned} \text{loss} &= \sum_{i=1}^m l_{\epsilon}(f(x_i), y_i) \\ \min_{w,b} & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m l_{\epsilon}(f(x_i), y_i) \end{aligned}$$

为保证所有训练数据都在间隔带内, 需要引入松弛变量 ξ_i 和 $\hat{\xi}_i$, 从而允许一些样本可以不在间隔带内。引入松弛变量的 SVR 模型可表示为:

$$\begin{aligned} \min_{w,b,\xi_i,\hat{\xi}_i} & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^m (\xi_i, \hat{\xi}_i) \\ \text{s.t.} & f(x_i) - y_i \leq \epsilon + \xi_i \\ & y_i - f(x_i) \leq \epsilon + \hat{\xi}_i \\ & \xi_i \geq 0, \hat{\xi}_i \geq 0, i=1,2,\dots,m. \end{aligned}$$

定义一个低维到高维的映射 ϕ , 并且引入核函数 $k(x, x_i)$ 将非线性回归问题转变为近似线性回归, 则 SVR 映入核函数后可重写为(2007) [18]:

$$\begin{aligned} f(x) &= w^T \phi(x) + b = \sum_{i=1}^m a_i y_i \phi(x_i)^T \phi(x) + b \\ &= \sum_{i=1}^m a_i y_i k(x, x_i) + b \end{aligned}$$

基于支持向量机回归对测试集数据进行预测, 模型的精度取决于正则化系数 C 、核函数 $k(x, x_i)$ 的选取以及容忍偏差 ϵ 大小。

SVR (支持向量回归)是一种回归模型, 它通过将样本点映射到高维特征空间中, 寻找最优超平面来进行回归分析。SVR 模型也可用于以下几个方面: 金融领域、医疗领域、工程领域、环境科学、数据挖掘等方面。

4. 实证分析

4.1. 数据源及其处理

研究中所涉及到的中国二氧化碳排放量数据主要来源于全球实时碳排放数据库(Carbon Monitor), 选取了共计 1551 条 2019 年 1 月 1 日至 2023 年 3 月 31 日的每日碳排放量并绘制如图 2 所示折线图。

通过观察趋势变化图可以认为碳排放量呈现出周期性的特点, 为了保证计算的有效性和模型可靠性, 本文首先利用归一化公式对数据进行放缩处理, 以消除量纲对于数据的影响。

$$X_i^* = \frac{X - X_{\min}}{X_{\max} - X_{\min}}$$

其中, X_i^* 为归一化后的值; X 为原始值; X_{\max} 和 X_{\min} 分别代表序列原始数据的最大值和最小值。其次

采用如图 2 所示的滑动窗口法对归一化数据集进行分割, 综合考虑数据集大小及碳排放受季节性影响, 本次滑动窗口大小 s 设置为 90。最后选择 75% 的数据作为训练集, 25% 的数据作为测试集分别利用 SARIMA、SVR 和 LSTM 模型对数据进行建模。整体研究思路如图 3 所示。

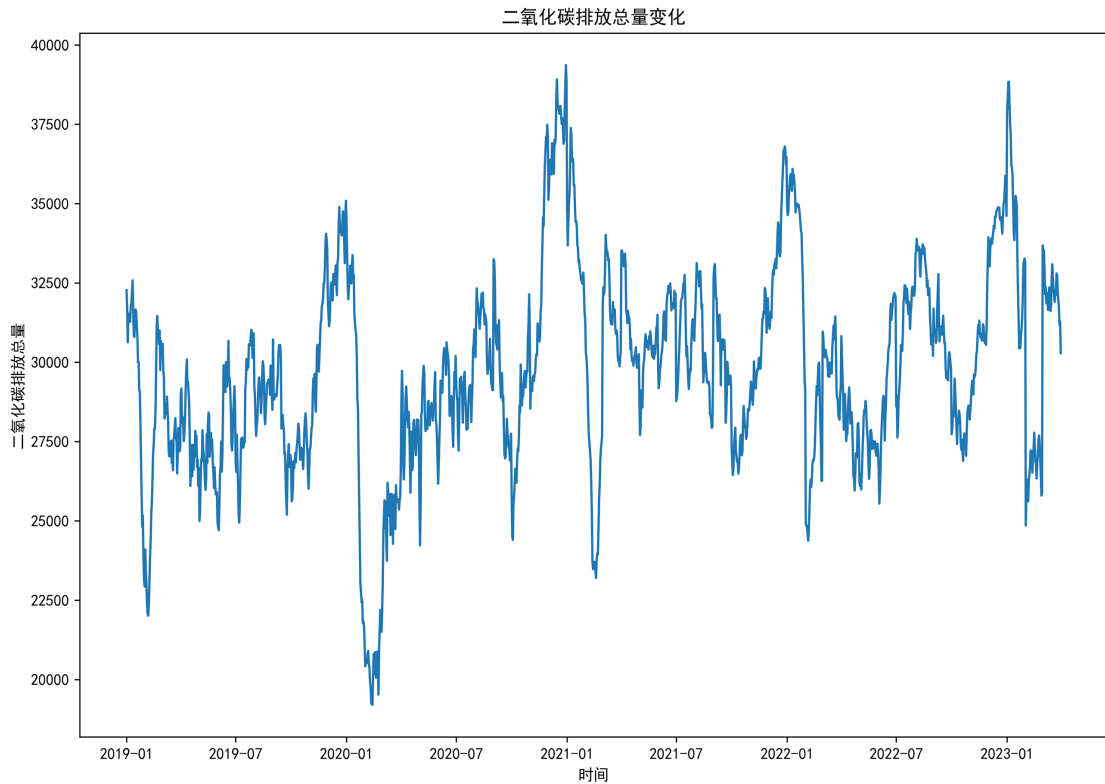


Figure 2. The trend change of national carbon emission data from January 2019 to March 2019
图 2. 2019.01~2023.03 全国碳排放数据趋势变化图

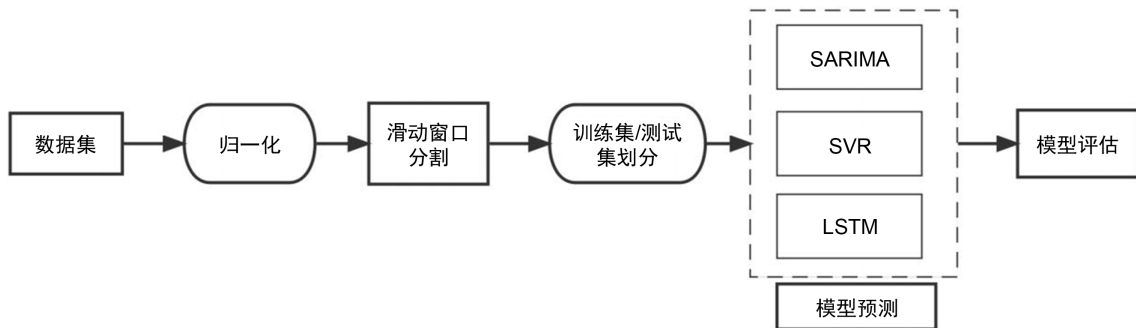


Figure 3. Overall design framework of carbon emission research
图 3. 碳排放研究总体设计框架

4.2. 模型参数选择

SARIMA 作为考虑周期性的 ARIMA 改进模型, 在利用其进行时间序列预测时, 首先需要对原始数据进行平稳性及白噪声检验。研究过程中首先对原时间序列绘制移动平均值和标准差, 并进行 ADF 检验, 根据对表 2 和图 4 结果进行分析可知时间序列具有平稳性。然后对数据进行白噪声检验, 结果显示 Q 统计量的 p 值小于 0.01, 所以在 0.01 的显著性水平下拒绝原假设, 可以认为序列是非白噪声序列。

Table 2. Results of unit root test of sequence X_t
表 2. 序列 X_t 的单位根检验结果

ADF 检验统计量值	p 值	1%水平下的临界值	5%水平下的临界值	10%水平下的临界值
-4.78	0.00	-3.43	-2.86	-2.57

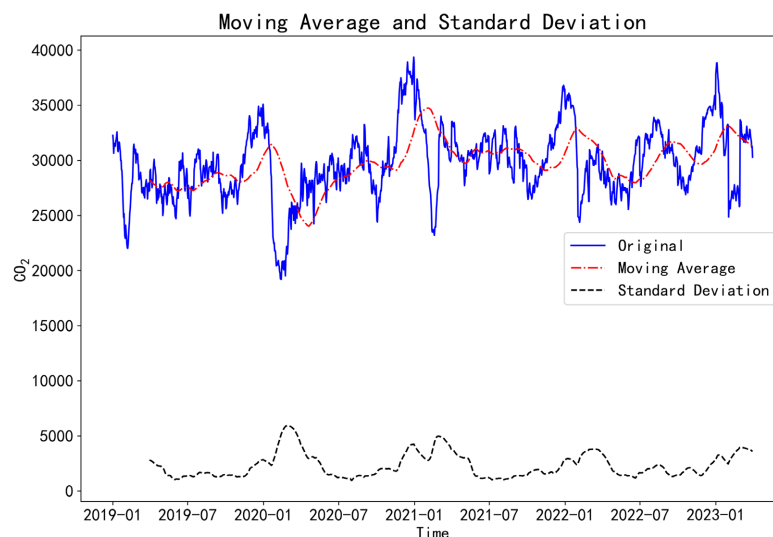


Figure 4. Moving average and standard deviation of sequence X_t
图 4. 序列 X_t 的移动平均值和标准差

对通过平稳性检验和白噪声检验的时间序列数据绘制自相关函数(ACF)图和偏自相关系数(PACF)图, 结果如图 5 所示, 通过对其进行分析, 可以初步确定 $p = 2$, $q = 2$ 。为精确获得 SARIMA 模型中参数 p , q , P , Q , 研究过程中利用网格调参(Grid Search CV)的方法以赤池信息准则(AIC)为判断标准选择 AIC 最小的参数作为模型的具体参数值。利用 python 计算可得模型最优参数为 SARIMA(2, 0, 2) \times (0, 0, 1, 90)。

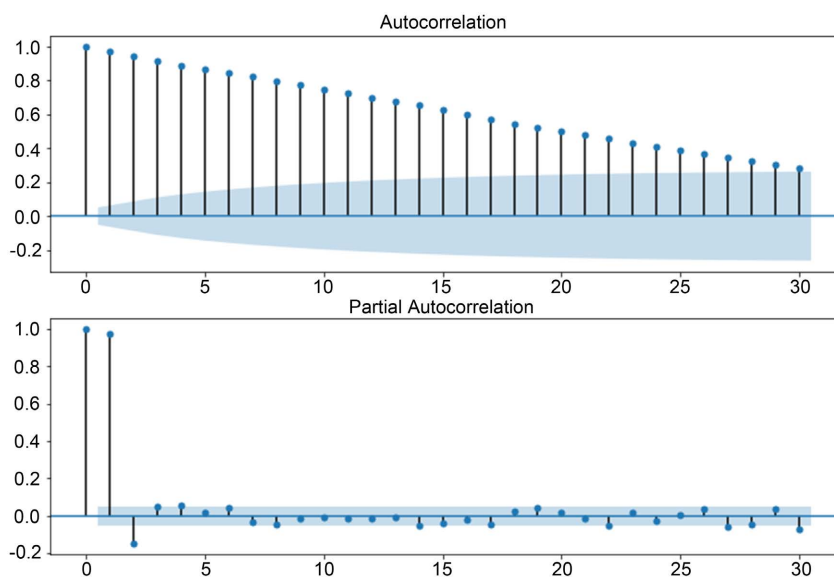


Figure 5. ACF and PACF plots of time series
图 5. 时间序列 ACF 和 PACF 图

该研究基于 tensorflow.keras 框架, 搭建 LSTM 长短时神经网络模型。模型采用单向编码方式, 所使用的 LSTM 网络模型超参数具体设置如下: 1 层输入层、2 层 LSTM 隐藏层以及两层全连接层, 其中每层隐藏层中隐藏神经元数量分别为 64 和 32。模型采用 Adam 优化算法寻找整体损失函数均方误差(MSE)的最优解, batch_size 大小设置为 3。如图 6 所示, 迭代次数的不同, 模型损失函数不同变化, 当迭代次数为 30 时, 训练集与测试集的损失函数均基本完全收敛, 故迭代次数设置为 30。模型最终参数如图 7 所示。

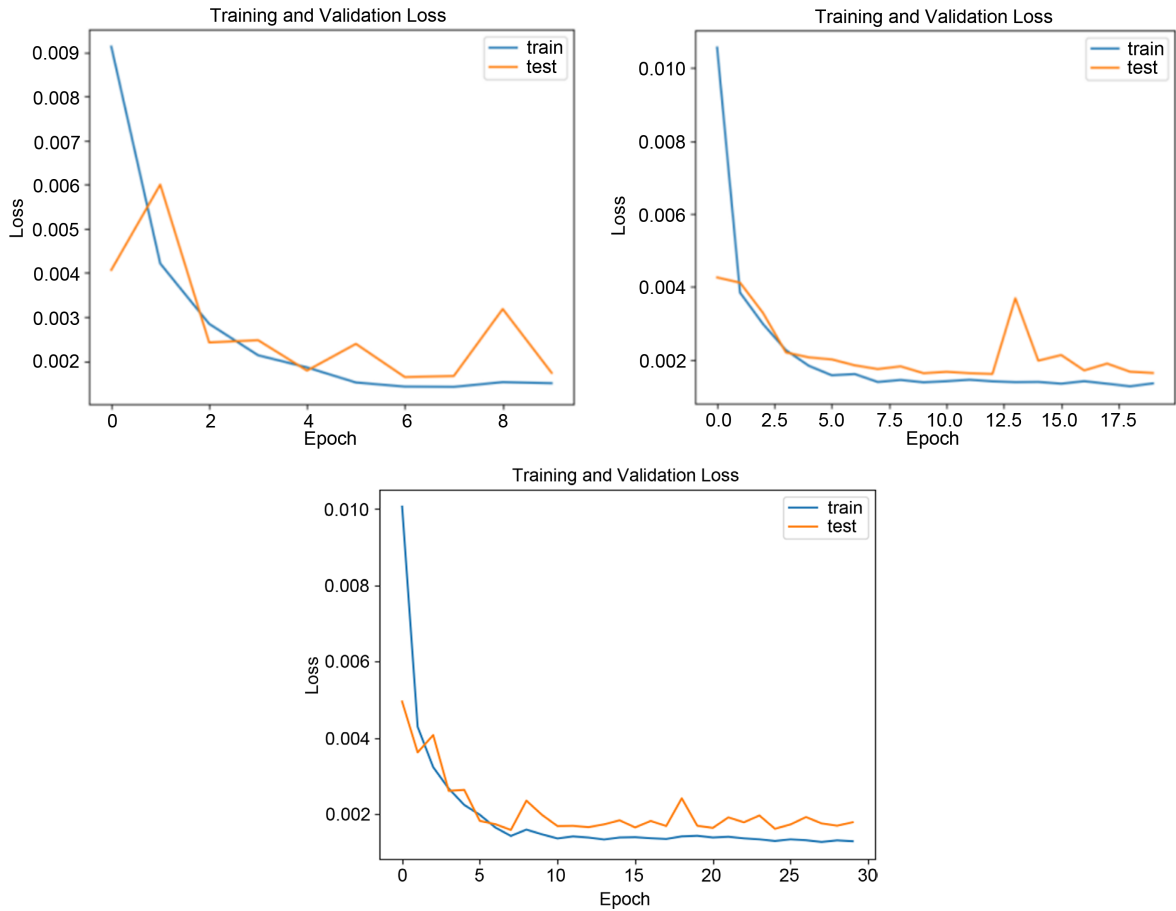


Figure 6. Trend chart of the number of iterations and loss function
图 6. 迭代次数与损失函数的变化趋势图

Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 90, 64)	16896
lstm_1 (LSTM)	(None, 32)	12416
dense (Dense)	(None, 2)	66
dense_1 (Dense)	(None, 1)	3

=====
 Total params: 29,381
 Trainable params: 29,381
 Non-trainable params: 0
 =====

Figure 7. Final parameters of LSTM neural network model
图 7. LSTM 神经网络模型最终参数

利用 SVR 对碳排放时间序列进行预测时, 核函数选择高斯核函数(RBF)对模型进行优化, 模型参数分别设置为 $C = 10$, $\epsilon = 0.05$, $\gamma = 0.5$ 。

4.3. 预测结果及其对比

4.3.1. 预测结果

按照训练集与测试集的划分比例, 并且基于以上确定的模型参数, SARIMA、LSTM、SVR 三个模型将在 2019 年 1 月至 2022 年 3 月的碳排放数据训练集上进行拟合, 接着在 2022 年 4 月至 2023 年 3 月的碳排放测试集上进行预测。各模型的预测效果如图 8 所示, 表 3 列举了最后 10 天的碳排放实际数据与不同模型的预测值对比情况。从图 8 可知 SVR 模型的预测效果相对较差, LSTM 和 SARIMA 模型对于测试集的预测效果相对较好, 从表 3 预测值的量化分析可知, SARIMA 模型的平均相对误差为 101.292, 相对其他两个模型来说误差更小。因此, 通过整体比较可知 SARIMA 模型具有更好的拟合预测效果。

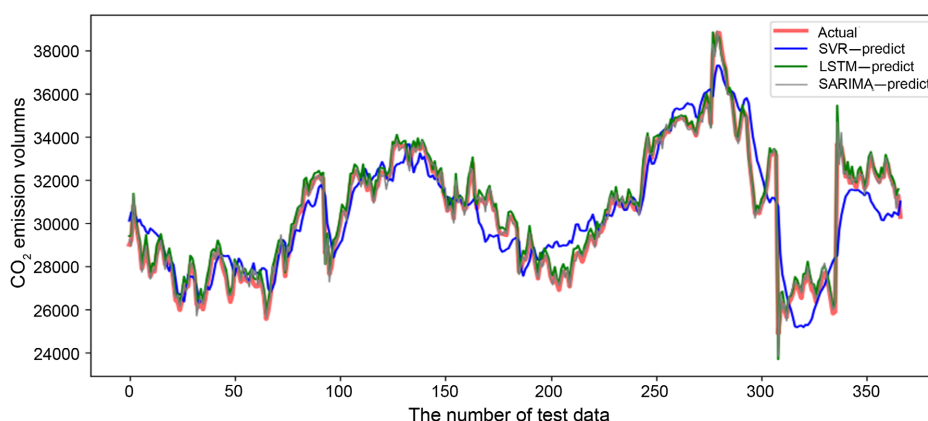


Figure 8. Comparison of carbon emission prediction results of different model test sets
图 8. 不同模型测试集碳排放量预测结果对比

Table 3. Comparison of experimental relative errors of predicted values of the test set in the last 10 days
表 3. 后 10 日测试集预测值对比实验相对误差

时间	真实值	SARIMA 模型		LSTM 模型		SVR 模型	
		预测值	误差	预测值	误差	预测值	误差
2023-3-22	32258.385	31833.436	-424.948	32111.883	-105.035	30158.913	-2099.472
2023-3-23	32119.902	32197.503	77.601	32589.463	464.778	30129.364	-1990.538
2023-3-24	32801.516	32120.169	-681.347	32274.021	-523.786	30327.631	-2473.885
2023-3-25	32746.971	33041.252	294.281	33176.867	359.064	30442.507	-2304.464
2023-3-26	32217.604	32703.437	485.833	32864.040	573.286	30380.413	-1837.191
2023-3-27	31969.388	32178.065	208.677	32268.588	292.755	30291.142	-1678.246
2023-3-28	31718.473	31715.898	-2.575	3216.275	456.209	30511.235	-1207.238
2023-3-29	31172.849	31772.034	599.185	31946.121	773.272	30447.653	-725.196
2023-3-30	31275.943	30682.072	-593.871	31910.61	117.043	30377.565	-898.378
2023-3-31	30284.347	31334.436	1050.089	31581.252	1342.989	31001.128	716.781
平均误差			101.292		375.0575		-1449.783

4.3.2. 模型对比

为了更为具体地比较模型之间的预测效果, 研究过程中选用平均绝对误差(MAE)、均方根误差(RMSE)以及判定系数(R^2)作为模型的评价指标, 衡量预测值与真实值之间存在的差异。各指标的具体计算公式为:

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

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2}$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

其中, y_i 为真实值; \hat{y}_i 为预测值; \bar{y} 为样本均值; n 为样本个数。

基于以上评价指标对不同模型的预测精度进行分析, 得到结果如表 4 所示。SVR 模型预测效果相对较差, MAE、RMSE 指标都远大于其余模型的预测精度, R^2 仅为 0.829 相对较小。SARIMA 模型与 LSTM 模型在预测精度上相差较小, 但总体来说, SARIMA 模型的 R^2 达到 0.912, 模型具有更好的预测效果, 预测误差更小。

Table 4. Comparison of prediction accuracy of different models

表 4. 不同模型的预测精度对比

指标	SARIMA	LSTM	SVR
MAE	479.89	489.79	883.82
RMSE	835.24	853.47	1161.21
R^2	0.912	0.908	0.829

5. 结语

本文使用 Python 软件对 2019 年 1 月至 2023 年 3 月的全国碳排放实时数据, 在考虑原时间序列具有趋势性的基础上, 分别建立了 SARIMA 模型、LSTM 模型以及 SVR 模型, 利用 MAE、RMSE 以及 R^2 三个指标分别对比分析了不同模型的预测精度。结果显示 SARIMA 模型对于具有趋势性的碳排放实时数据具有更好的预测效果, 而 SVR 模型对碳排放的拟合度和准确性相对较差。

参考文献

- [1] 冯相昭, 邹骥. 中国 CO₂ 排放趋势的经济分析[J]. 中国人口·资源与环境, 2008, 18(3): 43-47.
- [2] 袁路, 潘家华. Kaya 恒等式的碳排放驱动因素分解及其政策含义的局限性[J]. 气候变化研究进展, 2013, 9(3): 210-215.
- [3] Jiang, P., Gong, X.J., Yang, Y.R., et al. (2023) Research on Spatial and Temporal Differences of Carbon Emissions and Influencing Factors in Eight Economic Regions of China Based on LMDI Model. *Scientific Reports*, **13**, Article No. 7965. <https://doi.org/10.1038/s41598-023-35181-w>
- [4] 张矢宇, 杨杰, 田志武. 基于 LMDI 的交通运输碳排放因素分解及预测研究[J]. 武汉理工大学学报(信息与管理工程版), 2023, 45(2): 252-257.
- [5] 胡广阔, 李春梅, 惠树鹏. 基于改进 STRIPAT 模型在碳排放强度预测中的应用[J]. 统计与决策, 2016(3): 87-89.
- [6] 刘茂辉, 翟华欣, 刘胜楠, 岳亚云, 杨多堃, 李婧. 基于 LMDI 方法和 STIRPAT 模型的天津市碳排放量对比分析

- [J]. 环境工程技术学报, 2023, 13(1): 63-70.
- [7] 杨青, 彭若慧, 刘星星, 曹兰娟. 基于地理加权回归的省域碳排放影响因素研究[J]. 环境工程技术学报, 2023, 13(1): 54-62.
- [8] 肇晓楠, 谢新连, 赵瑞嘉. 基于滑动窗口动态输入 LSTM 网络的铁路运输系统碳排放量预测方法[J]. 交通信息与安全, 2023, 41(1): 169-178.
- [9] 池小波, 续泽晋, 贾新春, 张伟杰. 基于 WPD-ISSA-CA-CNN 模型的电厂碳排放预测[J]. 控制工程, 2023: 1-8.
- [10] Vahid, A. and Mark, G. (2023) Day Ahead Carbon Emission Forecasting of the Regional National Electricity Market Using Machine Learning Methods. *Engineering Applications of Artificial Intelligence*, **123**, Article ID: 106314. <https://doi.org/10.1016/j.engappai.2023.106314>
- [11] Zhao, J.J., Kou, L., Wang, H.T., He, X.Y., et al. (2022) Carbon Emission Prediction Model and Analysis in the Yellow River Basin Based on a Machine Learning Method. *Sustainability*, **14**, Article No. 6153. <https://doi.org/10.3390/su14106153>
- [12] Yuan, J.B., Liu, Y.S., Xu, J., Gao, S.H. and Li, X.T. (2019) Carbon Emission Prediction of Civil Buildings in China Based on Improved Grey Prediction Method. *Proceedings of 2019 2nd International Conference on Manufacturing Technology, Materials and Chemical Engineering (MTMCE 2019)*, Wuhan, 14-16 June 2019, 1065-1073.
- [13] Box, G.E.P., Jenkins, G.M., Reinsel, G.C. 时间序列分析: 预测与控制[M]. 北京: 中国统计出版社, 1997: 211-376.
- [14] Tian, C., Wang, H. and Luo, X. (2019) Time-Series Modelling and Forecasting of Hand, Foot and Mouth Disease Cases in China from 2008 to 2018. *Epidemiology & Infection*, **147**, e82. <https://doi.org/10.1017/S095026881800362X>
- [15] 王燕. 应用时间序列分析[M]. 北京: 中国人民大学出版社, 2016.
- [16] 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>
- [17] 杜树新, 吴铁军. 用于回归估计的支持向量机方法[J]. 系统仿真学报, 2003(11): 1580-1585, 1633.
- [18] Ustünu, B., Melssen, W.J. and Buydens, L.M.C. (2007) Visualisation and Interpretation of Support Vector Regression Models. *Analytica Chimica Acta*, **595**, 299-309. <https://doi.org/10.1016/j.aca.2007.03.023>