

## 基于多头注意力 CNN-LSTM 碳排放量预测研究

仵晓聪<sup>a</sup>, 冯鑫<sup>a</sup>, 蒋豪<sup>b</sup>

重庆工商大学 a. 机械工程学院智能装备绿色设计与制造重庆市重点实验室; b. 数学与统计学院, 重庆 400067

**摘要:**目的 由于碳排放具有不规则性和非线性的特点, 针对如何提升碳排放量预测效果, 提出一种基于多头注意力 CNN-LSTM 的碳排放预测方法。方法 基于 CNN 网络具有的空间特征提取优势和 LSTM 算法有效利用长、短期时间的依赖关系, 引入多头注意力机制以分配不同权重的方式, 采用多头注意力机制引导下的 CNN-LSTM 模型对我国碳排放量进行预测。结果 对过去 1 724 d 碳排放量的数据集进行训练和预测, 发现测试集的 RMSE、MASE、MAPE 指标数值分别为 0.080 2、0.302 0、0.105 7; 与 BP 神经网络算法、ARIMA(2, 1, 4)、GM(1, 1) 进行对比, 可以明显看出设计的模型在 3 种评价指标方面均优于其余预测算法。结论 采用基于多头注意力的 CNN-LSTM 碳排放预测方法能够有效提高预测精确度, 适用于时间序列预测。

**关键词:**碳排放量预测; 多头注意力机制; CNN-LSTM 模型; 多头注意力 CNN-LSTM 预测模型

**中图分类号:** TP183; X321 **文献标识码:** A **doi:** 10.16055/j.issn.1672-058X.2026.0003.014

### Research on Carbon Emission Prediction Based on Multi-Head Attention CNN-LSTM

WU Xiaocong<sup>a</sup>, FENG Xin<sup>a</sup>, JIANG Hao<sup>b</sup>

a. Chongqing Key Laboratory of Intelligent Equipment Green Design and Manufacturing, b. School of Mathematics and Statistics, Chongqing Technology and Business University, Chongqing 400067, China

**Abstract: Objective** Due to the irregularity and nonlinearity of carbon emissions, this study proposes a carbon emission prediction method based on multi-head attention CNN-LSTM to enhance the accuracy of carbon emission predictions.

**Methods** Leveraging the spatial feature extraction capabilities of CNN networks and the effective utilization of long-term and short-term dependencies by the LSTM algorithm, a multi-head attention mechanism was introduced to allocate different weights. A CNN-LSTM model guided by the multi-head attention mechanism was employed to predict carbon emissions in China.

**Results** The research selected a dataset of China's carbon emissions over the past 1, 724 days and used the model proposed in the paper to conduct training and prediction on this dataset. After testing, the values of the RMSE, MASE, and MAPE of the model on the test set were 0.080 2, 0.302 0, and 0.105 7, respectively. Comparing with the BP neural network algorithm, ARIMA (2, 1, 4), and GM (1, 1), the model designed in the paper outperformed the other prediction algorithms in terms of these three evaluation indicators. **Conclusion** The carbon emission prediction method based on multi-head attention CNN-LSTM can effectively improve the prediction accuracy and is suitable for time series prediction.

**收稿日期:** 2024-05-06 **修回日期:** 2024-08-20 **文章编号:** 1672-058X(2026)03-0116-09

**基金项目:** 重庆市自然科学基金项目资助(CSTB2022NSCQ-MSX0271); 重庆市高校创新研究群体项目资助(CXQT21024).

**作者简介:** 仵晓聪(2000—), 男, 四川成都人, 硕士研究生, 从事图像融合研究。

**通信作者:** 冯鑫(1982—), 男, 四川泸州人, 教授, 博士研究生, 从事光学图像处理、图像融合和深度学习研究。Email: 149495263@qq.com.

**引用格式:** 仵晓聪, 冯鑫, 蒋豪. 基于多头注意力 CNN-LSTM 碳排放量预测研究[J]. 重庆工商大学学报(自然科学版), 2026, 43(3): 116-124.

Wu Xiaocong, Feng Xin, Jiang Hao. Research on carbon emission prediction based on multi-head attention CNN-LSTM[J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2026, 43(3): 116-124.

**Keywords:** carbon emission prediction; multi-head attention mechanism; CNN-LSTM model; multi-head attention CNN-LSTM prediction model

实现“2030年碳达峰、2060年碳中和”,是以习近平总书记为核心的党中央统筹国内国际两个大局作出的重大战略决策,是构建人类命运共同体的庄严承诺<sup>[1]</sup>。预测我国碳排放量变化程度不仅可以评估当前政策制定的合理性,而且还能为今后科学合理制定减排政策提供决策依据和参考。因此,如何准确地预测碳排放量变得尤为重要。

目前,对于碳排放量的预测,我国学者研究成果主要分为两类:单一模型预测和组合类模型预测。在单一模型预测中,黄昕怡等<sup>[2]</sup>利用灰度模型 GM(1,1) (Grey Model(1,1)),预测江苏省碳排放量,通过残差检验法得到该方案的可行性,该模型利用江苏省2000—2004年数据进行预测,获得了不错的效果;周树涛等<sup>[3]</sup>利用自回归移动平均(Autoregressive Integrated Moving Average, ARIMA)模型进行预测,采用我国1997—2019年数据,通过平稳性验证和模型阶数确定等步骤,确定了该模型适用于我国碳排放预测,并给出了减排建议;赵雄飞等<sup>[4]</sup>通过引入长短期记忆网络(Long Short-Term Memory, LSTM),并结合我国1978—2016年数据进行预测,分析预测结果,发现科技进步、产业结构和能源结构等因素会对我国碳排放量造成影响。其他的预测方法还有对数平均迪维西亚指数模型(Logarithmic Mean Divisia Index model, LMDI)等<sup>[5]</sup>。以上模型预测结果较为准确,但是所采用数据量较小,容易出现欠拟合情形,一些模型可能需要大量的参数调整和优化,增加了模型使用的复杂性,且预测时未能充分捕捉数据复杂性和变化趋势,导致预测有偏差。在组合类模型预测中,张学清等<sup>[6]</sup>提出用 CNN-LSTM 模型预测中国碳排放量,该模型解决了卷积神经网络(Convolutional Neural Networks, CNN)难以记忆早期时间序列信息的问题,也克服了 LSTM 模型对于预测数据精度不够的情况,利用700多天的碳排放量数据进行训练和测试,发现实验结果具有较小的误差,更高的精度;楼泽瑶等<sup>[7]</sup>采用加权延时并行势场与回归(Weighted Time-Delayed Parallel Potential Field and Regression, WTD-PPAR)方法,可以很好地解决碳排放数据存在非线性及波动性问题,然后在 PPAR 模型上预测出我国在2029年达到碳排放峰值;王文佳等<sup>[8]</sup>采用灰狼优化-长短期记忆网络(Grey Wolf Optimizer-

Long Short-Term Memory, GWO-LSTM)预测方法,先引入 GWO 算法得到 LSTM 模型最佳参数,然后再利用最佳参数构建 LSTM 模型,结果表明该模型在辽宁省工业碳排放预测方面有较高的精确度。

以上模型预测将两个或多个模型优势结合,相比于传统统计模型和单一的深度学习模型精确度更好,同时预测时限更长,这对我国实现碳达峰碳中和目标具有指导意义。尽管预测结果较好,但是模型并不具备良好的鲁棒性,容易出现过拟合情形。此外,由于碳排放容易受到额外因素影响,长期预测中可能存在一定的收敛性问题,导致预测值波动相对平稳的趋势不够明显。由于数据训练和预测需要大量的计算资源和时间,这可能限制了其在实时预测或资源受限的应用场景中使用。在国外学者中,文献[9]采用基回归随机影响模型对撒哈拉沙漠以南和东南亚国家二氧化碳排放量进行分析,发现农业、林业、渔业和人类发展指数对撒哈拉以南非洲和东南亚国家的碳排放量具有显著的影响力;文献[10]采用自回归结构方程模型对泰国二氧化碳排放量预测,结果表明在接下来的15年内,排放量将持续增长,会超过预定上限,并对此给出了一些建议。其他预测模型还有文献[11]等。以上算法从影响因素角度进行,对诸多碳排放量的影响因素进行权重分析,找到了影响碳排放的主要因素。但由于因素影响不确定,收集的数据不够完整,因此只能在特定因素数据下分析,无法全面反映实际情况,导致预测结果出现误差。

尽管以上算法在进行影响因素分析时可能存在关注度不准确或权重自动调整的问题,但是注意力机制的出现带来了新的解决方案。注意力机制最早于计算机视觉领域出现<sup>[12]</sup>,它模仿人类视觉系统的工作方式,使神经网络能够在处理输入数据时,集中注意力于重要的区域或特征,通过自主针对性地选择信息,从而提高处理效率和性能。因此将注意力机制与其他网络模型结合有助于解决一些传统方式难以处理的问题<sup>[13]</sup>。多头注意力机制是在自注意力机制的基础上发展起来的,是自注意力机制的变体,旨在增强模型的表达能力和泛化能力。它通过使用多个独立的注意力头,分别计算注意力权重,并将它们的结果进行拼接或加权求和,从而获得更丰富的表示。为了提高碳排放量预测

的精准度和可靠性,本文研究引入多头注意力机制的表达和泛化能力,并将 CNN 模型的空间特征提取和 LSTM 处理时间序列的优越性结合,提出多头注意力 CNN-LSTM 预测模型。该模型先采用 CNN 模型进行初步特征提取,再用 LSTM 处理时间序列,并保留长期依赖性,之后再引入多头注意力机制,进行并行运算,这样可以学习到时间序列中不同位置之间的不同依赖关系,大大提高模型的运算能力。模型能够同时学习和捕捉到不同位置之间的依赖关系,从而提高对序列数据的理解能力。结合 CNN、LSTM 和多头注意力机制三者优势之后,模型能够更准确地预测序列数据。尽管模型复杂度较高,但通过优化网络结构和参数调整提高了模型的泛化能力,使其能够更好地适应不同数据集和不同问题的需求。实验表明:本文提出的方法具有较好的预测性能和泛化能力,能够使预测结果更接近真实值,可以为国家制定针对性的政策提供参考。

## 1 多头注意力 CNN-LSTM 预测模型

### 1.1 模型提出动机

卷积神经网络通过层层堆叠卷积层、池化层、填充层等步骤,可以逐步提取输入数据的抽象特征<sup>[14]</sup>。通过局部连接和权值共享机制,能够有效处理大量数据。将输入数据提取出显著特征,过滤掉其余不显著信息,从而减少参数数量和计算复杂度,提高模型的泛化能力。由于在卷积神经网络中可能出现梯度消失或爆炸问题,于是引入 LSTM。LSTM 由 RNN 改进而来,在原基础上加入了细胞态,同时将各门之间运算复杂化,使

得该网络结构能够学习并记忆长期依赖关系,同时可以有效缓解梯度消失或爆炸的问题<sup>[15]</sup>。但是,传统 CNN 和 LSTM 模型可能会存在一些缺陷。CNN 无法处理复杂时间序列,同时在面对特征多的对象时容易形成数据灾难。LSTM 模型在处理很长的时间序列时,仍然可能会出现梯度爆炸等问题,同时由于计算时间跨度较大加上涉及参数较多,会出现计算时间大大增加的情形。

注意力机制是一种用于选择关键信息的机制,它模拟了人对不同信息施加不同的关注度,从而更加有效地选择有用的信息。基于以上情况,在“多头注意力机制的 CNN-LSTM 模型”中,CNN 优势在于提取数据空间特征,LSTM 优势在于有效处理以往时间数据,给出预测数据。加入多头注意力机制可以计算不同时刻对输出数据的影响,改善 LSTM 因数据太多而遗忘重要信息的情况,以概率的方式分配权重,使得模型能够找到隐藏在历史数据中的特征并且给予不同的关注度<sup>[16-17]</sup>。

### 1.2 模型结构

基于以上分析,本文提出的模型整体架构如图 1 所示。首先,将输入数据经过 CNN 网络,并重新塑造维度,然后再将数据传入 LSTM 网络,输出结果将分别作为查询、键和值传递给注意力层。每个头都会学习到不同的注意力分布,在计算注意力时,通过计算查询和键之间的点积,然后使用 Softmax 函数对注意力权重进行归一化,从而确定每个位置对其他位置的关注度。再经过全连接层输出结果。该模型框架旨在优化信息处理效率和结果的准确性。

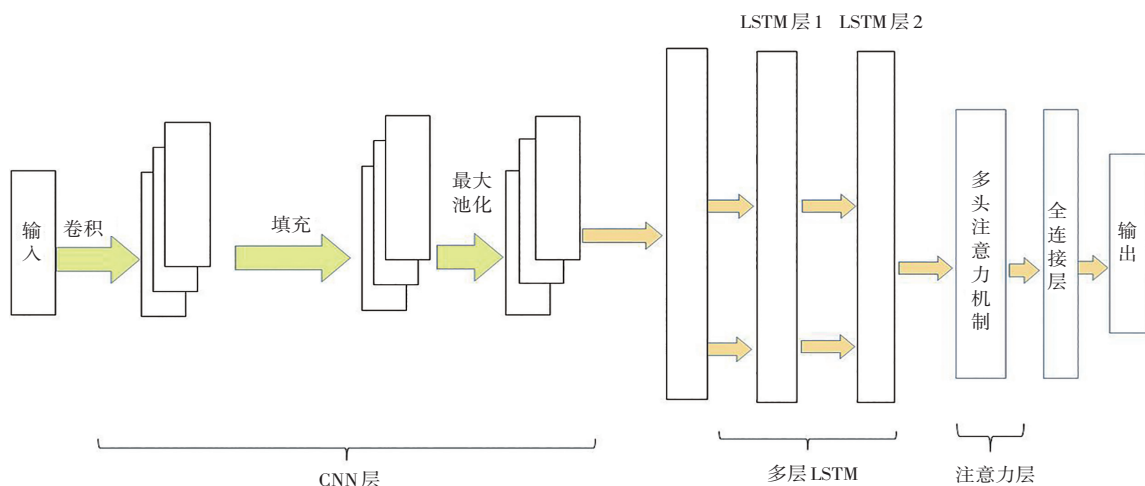


图 1 多头注意力 CNN-LSTM 模型结构

Fig. 1 Structure of multi-head attention CNN-LSTM model

首先,输入数据格式为[批量大小,时间步数,特征维度],设批量大小为 1,时间步数为  $t$ ,特征维度为  $n$ ,则输入序列可以实数矩阵  $\mathbf{R}^{t \times n}$ ,设  $\mathbf{x}_j$  为  $\mathbf{R}^{t \times n}$  中第  $j$  个时间步数长度的向量。将数据输入模型进行训练。

#### 1.2.1 预测模型特征提取层设计

CNN 层可以处理数据空间上的特征联系,弥补 LSTM 无法捕捉到的数据在空间上的特点<sup>[18]</sup>。CNN 通过稀疏连接和共享权值的方式,能够有效地从原始数

据中提取特征信息,卷积层通过滑动卷积的方式在输入数据中筛选局部信息,然后通过池化层对特征信息进行降维,再利用填充层调整输出信息大小以确保输入和输出均保持相同尺寸<sup>[19]</sup>。经过卷积层将特征放大,再经过最大池化层缩小数据信息,过滤作用不太大的信息。由于本文是针对一维时间序列,所以本模型采用一维卷积公式。设卷积核数目为  $m$ ,长度为  $L$ ,则  $x_{j,j+L-1}$  表示为  $\mathbf{R}^{n \times n}$  中第  $j$  个时间步长到第  $j+L-1$  个时间步长的实数矩阵。假设该卷积核滑动步长为 1,权重  $\omega_a$  是一个  $n * L$  的实数矩阵。对每个  $L$  时间步长序列做卷积运算,计算结果为

$$y_i = f(\omega_a \otimes x_{j,j+L-1} + b_i) \quad (1)$$

式(1)中,  $f$  表示一个非线性激活函数,本模型使用的激活函数是 ReLU 函数;  $b_i \in \mathbf{R}$ , 表示某个偏置项。卷积之后还需要再加入 padding =  $p$ , 即在卷积之后的序列长度进行填充  $p$  个零值。这样便可以使得输出与输入序列尺寸相同,原因是卷积操作通常会使得特征图尺寸缩小,得到新的特征长度为  $t+2p-L+1$ , 即  $y = [y_1, y_2, \dots, y_{t+2p-L+1}]^T$ 。由于卷积层一共有  $m$  个卷积核,因此会产生  $m$  个特征结果。在填充之后,进入最大池化操作。设池化尺寸为  $L_1$ , 池化时的滑动步长也为  $L_1$ , 这样设计可以确保不会有重叠的池化区域。池化之后可以得到长度为  $L^* = \lfloor \frac{t+2p-L+1-L_1}{L_1} \rfloor + 1$ , 表示输出长度等于将输入长度减去核长度再加一,再除以步长并向下取整,最后再加 1。其计算公式为

$$\hat{y} = \max(y_j, y_{j+1}) \quad (j=1, 2, \dots, L^*) \quad (2)$$

式(2)中,  $\hat{y}$  表示序列中较大的数值,长度为  $L^*$  的向量。

### 1.2.2 时间序列分析层设计

LSTM 主要用于解决一般循环神经网络 (Recurrent Neural Network, RNN) 存在的不能记住早期时间序列信息的问题<sup>[20]</sup>。比较起 RNN, LSTM 额外加入了候选态、细胞态和隐状态。隐状态  $h_t$  主要储存近期记忆,细胞态  $c_t$  储存长期记忆,可以缓解由梯度消失带来的损失,隐藏储存短期记忆。本模型采用多层 LSTM, 在计算时,上一 LSTM 层的输出是下一 LSTM 层的输入,在一次又一次传递之后,最后的 LSTM 输出进入多头注意力层中。

### 1.2.3 多头注意力机制层设计

多头注意力层通过引入多个并行的注意力头,允许模型在处理输入序列时从多个维度学习信息<sup>[21]</sup>。每个注意力头独立地关注不同的方面,并采用各自的权重分配方式,使得整体模型能够综合多种视角和特征,从而提供更为丰富和多样的数据表征。多头注意力能

够分配 LSTM 中重要的时间步权重,从而有效提高模型的预测精准度。设模型权值为  $v = [v_1, v_2, \dots, v_i]$ , 键值为  $k = [k_1, k_2, \dots, k_i]$ , 查询值为  $q = [q_1, q_2, \dots, q_i]$ 。设权值  $v$  对应的权重为  $\omega^v$ , 设键值  $k$  对应的权重为  $\omega^k$ , 查询值  $q$  对应的权重为  $\omega^q$ ,  $\omega^v, \omega^k, \omega^q$  由大量数据训练得到。将 LSTM 层输出向量分别作为多头注意力层的输入,即  $k, q, v$ ; 再使用 Softmax 函数进行归一化,得到每一个隐藏层向量的分配权重,权重大小表示每个时间步的隐状态对于预测结果的重要程度。权重计算公式为

$$Q = \omega^q q \quad (3)$$

式(3)中,  $Q$  表示查询值  $q$  所对应的权重,  $\omega^q$  表示查询值  $q$  所对应的权重。

$$K = \omega^k k \quad (4)$$

式(4)中,  $K$  表示键值  $k$  所对应的权重,  $\omega^k$  表示键值  $k$  所对应的权重。

$$V = \omega^v v \quad (5)$$

式(5)中,  $V$  表示权值  $v$  所对应的权重,  $\omega^v$  表示权值  $v$  所对应的权重。

$$\beta = \text{Softmax} \left( \left( \frac{QK^T}{\sqrt{d_k}} \right) V \right) \quad (6)$$

$$C = \sum_{i=1}^{L^*} \beta h \quad (7)$$

式(6)中,  $\beta$  表示再经过 Softmax 激活函数之后的向量,  $\sqrt{d_k}$  表示缩放因子。式(7)中,  $h$  表示最后一层 LSTM 的隐藏层输出结果,  $C$  表示加权求和之后的结果。

注意力机制决定输出权重,先通过一次线性层,接着进入激活函数,再次通过线性层将预测结果输出。

### 1.3 模型训练

多头注意力 CNN-LSTM 模型的训练过程如图 2 所示。

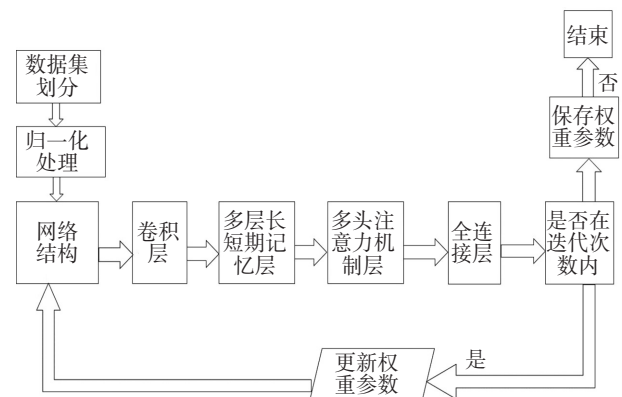


图 2 多头注意力 CNN-LSTM 模型的训练过程  
Fig. 2 Training process of multi-head attention CNN-LSTM model

(1) 函数通过遍历给定的数据,使用一个滑动窗口方法来准备数据。对于每个索引,从当前索引开始,按照特定步长选择每个时间步长的第一个特征数据作为一组特征,以将数据准备成适用于 LSTM 模型;训练集和测试集划分比例为 8 : 2,即 80%数据用于训练,20%用于测试。所有训练集与测试集数据都经过预处理并改变形状符合卷积层输入:

$$x'_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (8)$$

式(8)中, $x_i$ 表示原序列, $x_{\min}$ 表示原序列中的最小值, $x_{\max}$ 表示原序列中的最大值, $x'_i$ 表示原数据经过归一化之后的值。

(2) 在进行卷积运算之前对输入张量进行转置操作,将 feature\_size 和 time\_step 两个维度进行交换。其中,feature\_size 表示每个步长对应的特征数量,time\_step 表示时间步长。在转置之后,通过卷积操作处理一维张量  $x$ ,再进行填充。然后通过最大池化操作对卷积结果  $x$  进行池化。接着,代码获取批次大小,即输入张量的第一个维度大小。 $x$  的形状是 [batch\_size, feature\_size, time\_step], batch\_size 表示批次大小。因此,卷积运算应用于每个批次的每个特征,而时间步则视为输入序列中的每个点。由于 LSTM 模型接收的输入形状是 [batch\_size, time\_step, input\_size], input\_size 表示输入大小,所以将上一步卷积操作后的输出张量进行转置,形状变为 [batch\_size, feature\_size, time\_step],与输入张量  $x$  的形状相同。

(3) 在 LSTM 层对输入张量进行处理。LSTM 模型在每个时间步上接收输入,并输出当前时间步的隐藏状态 output,以及最终的隐藏状态 (h\_0, c\_0)。(h\_0, c\_0) 分别代表 LSTM 模型的初始隐藏状态和细胞状态。输出的 output 形状为 [batch\_size, time\_step, hidden\_size], 其中 hidden\_size 是 LSTM 隐藏状态的维度。

(4) 注意力机制层用来计算注意力加权的输出。注意力机制通常用于加强模型对输入序列中不同时间步的关注程度,从而提高模型的性能。这里采用 dropout 为 0.8,即丢弃一部分注意力值的概率为 0.8,即在每次训练迭代中,有 80%的概率会将某些神经元的输出置为 0。这样能够有效减少模型过拟合,提高其泛化能力,同时控制模型的复杂度。

(5) 在全连接层中,将注意力加权的输出张量进

行展开,将其从 3 维张量展平成 2 维张量。这一步通常是为了将数据准备成全连接层的输入形式。对展开后的输出进行全连接层计算、ReLU 激活函数操作,最后输出模型的预测结果。

(6) 判断是否在迭代次数内。倘若在迭代次数内,再次进行训练,更新权重参数,直到训练集损失函数呈现收敛形式;倘若不在迭代次数内,保存训练集得到的训练权重参数,输出之后进行还原,以便观察模型预测值。

$$x_i = x'_i (x_{\max} - x_{\min}) + x_{\min} \quad (9)$$

式(9)中, $x_i$ 表示还原之后的序列, $x_{\min}$ 表示原序列中的最小值, $x_{\max}$ 表示原序列中的最大值, $x'_i$ 表示原数据经过归一化之后的值。

#### 1.4 模型测试过程

多头注意力机制的 CNN-LSTM 模型训练过程如图 3 所示。先载入在训练过程中得到的权重,将测试集输入进模型中,得到预测之后的结果;然后计算给定的 3 个评价指标;接着将输出之后的结果反归一化,恢复到归一化之前的数量级。为便于观测,绘制归一化之后的预测结果。

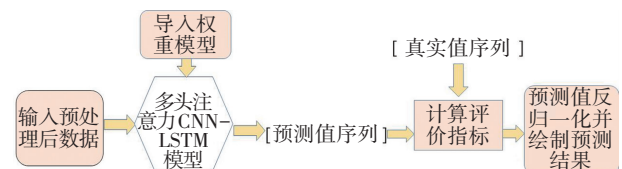


图 3 多头注意力机制的 CNN-LSTM 模型的测试过程

Fig. 3 Testing process of CNN-LSTM model with multi-head attention mechanism

## 2 仿真实验与结果分析

### 2.1 实验数据

本文采用的实验数据来自 CarbonMonitor。Carbon Monitor 提供了以日为分辨率的全球二氧化碳排放量数据,为科学研究和政策评估提供了重要的基础数据支持。本实验选取我国碳排放量从 2019-01-01—2023-09-30,共计 1 724 d。对数据集进行划分,训练集为 80%,共 1 371 条数据,测试集 20%,共 343 条数据。

### 2.2 实验平台与实验参数

本文算法实验的硬件配置:GPU 配置为 RTX3080,内存为 31 G。软件平台为 CUDA Version = 11.4,Python 版本为 3.12.2,Pytorch 版本为 2.2.1+cu118。

根据多次调试之后设置的参数如下:time\_step(时间

步长)为 20, batch\_size(批次大小)为 256, hidden\_size(隐藏层)为 128, learning\_rate(学习率)为 0.000 5, num\_layer(LSTM 层数)为 1, epochs(训练轮次)为 15, CNN\_out\_channels(CNN 输出通道)为 25, Multi\_heads\_num(多头注意力头数量)为 16。

### 2.3 评价指标

为了更好地客观评价模型预测效果和实际应用情况,本文引用 3 个相关评价指标来衡量多头注意力 CNN-LSTM 模型的优越性。指标分别为预测缩放误差平均值、均方误差和平均绝对百分比误差。3 个评价指标如下:

(1) 预测缩放误差平均值<sup>[22]</sup> MASE (Mean Absolute Scaled Error)。该指标是一种用于评估时间序列预测模型准确性的指标,可以消除因数据缩放引起的误差比较问题。数值越低,表明预测效果越好。其公式为

$$\mu_{\text{MASE}} = \frac{1}{n} \sum_{i=1}^n |q(t)| \quad (10)$$

$$q(t) = \frac{V(t) - p(t)}{\sum_{i=1}^n \frac{|V(i) - V(i-1)|}{n-1}} \quad (11)$$

式(10)中,  $V(t)$  为  $t$  时刻的序列真实值,式(11)中  $p(t)$  表示  $t$  时刻的预测值。

(2) 均方误差 RMSE(Root Mean Square Error)<sup>[23]</sup>。均方误差是回归损失函数中最常用的误差,它是预测值  $\hat{y}_i$  与目标值  $y_i$  之间差值平方和的均值再开方。数值越低,表明预测效果越贴近真实值,该预测结果越精确。其公式为

$$\mu_{\text{RMSE}} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (12)$$

式(12)中,  $y_i$  表示真实值,  $\hat{y}_i$  表示预测值,  $n$  表示每个样本的数据个数。

(3) 平均绝对百分比误差 MAPE (Mean Absolute Percentage Error)<sup>[24]</sup>。该指标可以反映预测值偏离真实值的百分比程度,不会受到量纲影响。数值越低,表明预测效果越好。其公式为

$$\mu_{\text{MAPE}} = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \quad (13)$$

式(13)中,  $y_i$  表示真实值,  $\hat{y}_i$  表示预测值,  $n$  表示每个样本的数据个数。

### 2.4 对比实验

为了验证本文提出的多头注意力 CNN-LSTM 模型的有效性,这里将其与深度学习中常用预测算法之一的 BP 神经网络算法<sup>[25]</sup>,经典时间序列预测算法中的 ARIMA(2,1,4)<sup>[3]</sup>、GM(1,1)<sup>[2]</sup>和多头注意力 CNN-LSTM 模型进行对比。为了统一变量,所有算法得到的预测结果均归一化处理之后再行客观指标评价,其结果如表 1 所示。

表 1 不同方案预测结果对比

Table 1 Comparison of prediction results for different schemes

预测方法	$\mu_{\text{RMSE}}$	$\mu_{\text{MASE}}$	$\mu_{\text{MAPE}}$
BP 神经网络算法	0.082 4	0.379 0	0.117 0
GM(1,1)	0.345 8	8.576 7	0.605 2
ARIMA(2,1,4)	0.465 0	11.535 6	0.886 7
多头注意力 CNN-LSTM 模型	<b>0.080 2</b>	<b>0.302 0</b>	<b>0.105 7</b>

由表 1 的 4 种预测方法对比结果可以看出:多头注意力 CNN-LSTM 预测模型比 BP 模型在各指标上均略有提升;比 GM(1,1)和 ARIMA(2,1,4)在各指标上均有显著提升。说明多头注意力 CNN-LSTM 模型充分利用了三者特点,即利用 CNN 空间特征提取、LSTM 时间特征提取和多头注意力机制对不同时间特征的重要性进行了合理分配。这种结合方式使得多头注意力 CNN-LSTM 模型预测更加精确,提高了其在时间序列预测任务中的性能和泛化能力。4 种预测算法结果可视化如图 4 所示。

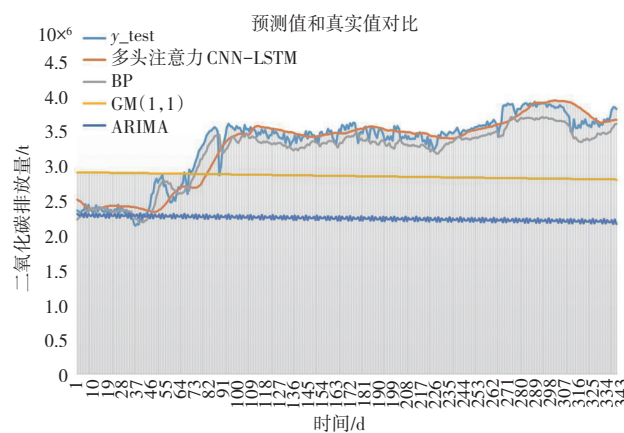


图 4 预测值和真实值对比

Fig. 4 Comparison between predicted and actual values in comparative experiments

图 4 为各预测方法的预测值和原始数据真实值 ( $y_{\text{test}}$ ) 对比结果图。可以观察到 ARIMA(2,1,4)方

法出现了最大误差,说明 ARIMA 算法并不适合于进行短期预测,只能用于近期预测;GM(1,1)方法预测效果不佳,出现了较大误差,呈现出一条直线,说明学习效果并不优秀;BP 模型预测值与真实值虽然变化趋势大体相同,但是却存在一定差距;而本文提出的基于多头注意力 CNN-LSTM 的预测模型预测值与真实值重合部分较多,损失也相对较小,说明本文所提预测方案有着更好的预测效果。

### 2.5 消融实验

为了进一步分析所提模型在预测中的效果,本文设计了相应的消融实验,包括 LSTM、CNN-LSTM 和提出的多头注意力 CNN-LSTM 模型。所有公共超参数设置相同,仅每个模型各自具有的超参数不同。CNN-LSTM 模型中增加 CNN 输出通道数为 25,多头注意力 CNN-LSTM 模型中加入 CNN 输出通道数为 25,注意力机制头数为 16。

当模型经过 15 次训练迭代后,训练集上的损失函数已经在某一数值附近波动且收敛,这表明模型在训练数据上的参数调整趋于稳定,已基本完成学习过程,如图 5 所示。为了评估这些模型的性能,进行了多次测试,测试结果如表 2 所示。

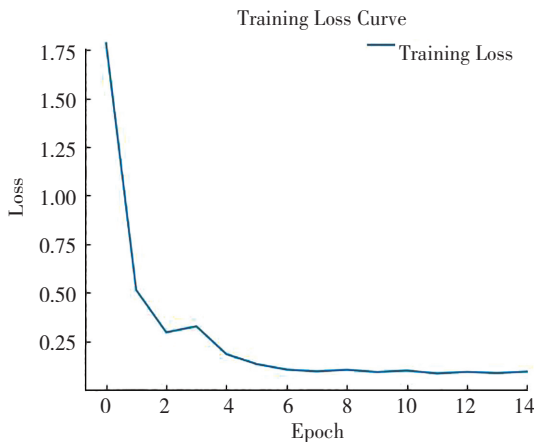


图 5 训练集损失函数变化

Fig. 5 Variation of the training set loss function

表 2 消融实验各模型评价结果

Table 2 Evaluation results of various models in ablation experiments

预测方法	$\mu_{RMSE}$	$\mu_{MASE}$	$\mu_{MAPE}$
LSTM 模型	0.137 5	0.635 7	0.258 3
CNN-LSTM 模型	0.091 0	0.383 0	0.141 0
多头注意力 CNN-LSTM 模型	<b>0.080 2</b>	<b>0.302 0</b>	<b>0.105 7</b>

从表 2 数据可以明显看出:多头注意力 CNN-LSTM 模型相对于传统 LSTM 模型在预测性能上有显著提升,这得益于其引入了卷积神经网络和多头注意力机制。通过卷积神经网络,模型能够更有效地学习时间序列中的局部特征和模式,而多头注意力机制则允许模型在不同时间步上分配不同的注意力权重,从而更好地捕捉序列中的重要信息。这些改进使得多头注意力 CNN-LSTM 模型在 MSE、MASE 和 MAPE 指标上都有显著的提升,表明其在时间序列预测任务中具有更好的性能和泛化能力。消融实验在测试集上的折线对比如图 6 所示。

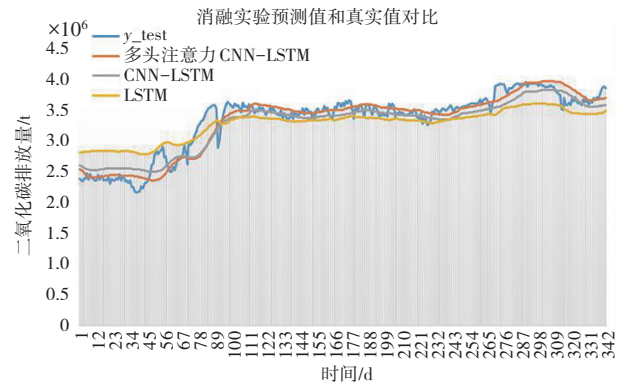


图 6 消融实验预测值和真实值对比

Fig. 6 Comparison between predicted and true values in ablation experiments

图 6 为本文提出的多头注意力 CNN-LSTM 预测模型各模块的消融实验。可以明显看出:多头注意力 CNN-LSTM 模型与 CNN-LSTM 预测趋势和真实值波动趋势几乎相同,在调节参数过程中发现多头注意力 CNN-LSTM 模型不容易陷入欠拟合,具有更好的泛化性。对比 LSTM 模型,多头注意力 CNN-LSTM 模型与预测值的重合范围更广,且损失结果更小,显示出更精确的预测能力。此外,在调节参数时发现该模型具有更加优良的鲁棒性。由此可以得出:多头注意力 CNN-LSTM 模型对时间序列类型的预测问题能够作出更准确的预测。

### 2.6 结果分析

通过对比实验与消融实验,可以发现多头注意力 CNN-LSTM 模型在 3 个指标上的表现均优于其他预测方法。原因是多头注意力机制的加入为 CNN-LSTM 模型带来了更强的表达能力和更灵活的特征学习能力。这种机制允许模型同时关注时间序列中的不同位置信

息,并能够在不同时间步上分配不同的注意力权重。相比之下,LSTM模型、CNN-LSTM模型可能会受限于固定的卷积核大小和全局的注意力分配,无法有效捕捉时间序列中的局部特征和长期依赖关系。而传统机器学习由于无法通过学习数据中的模式自动调整权重参数,这大大降低了学习效率。因此,加入多头注意力机制后,模型能够更好地适应不同长度和复杂度的时间序列数据,从而提高了预测性能和泛化能力。由此可以说明本文所提出的多头注意力 CNN-LSTM 模型更适用于时间序列预测。

### 3 结 论

精确预测碳排放量对于我国制定针对性的环境保护政策具有重要意义。本文在 CNN 网络模型与 LSTM 模型基础之上,引入多头注意力机制,提出基于多头注意力 CNN-LSTM 的碳排放预测网络。通过卷积神经网络中的卷积层和池化层提取局部特征,不仅减少了参数量和计算复杂度,而且还提高了模型的泛化能力。利用 LSTM 具有的对过去时间的敏感性,能够学习到时间序列中的特征,解决容易出现的梯度消失问题。而多头注意力机制通过多个独立注意力机制并行运算,起到了一个集成作用,有助于防止模型过拟合。基于以上优点,多头注意力机制的 CNN-LSTM 模型与其他模型相比较,在测试集上表现更为出色,同时泛化性和鲁棒性也更优良,证明了本模型可以用于时间序列预测。

本文方法虽然已经取得了显著效果,但是仍存在不能完美拟合真实值的问题。未来研究可以进一步优化网络结构,例如 CNN、LSTM 和 ATTENTION 3 种网络结构交替出现以增强模型拟合能力。这种方法的深入探索将有助于进一步提升该算法的预测精确性。

#### 参考文献(References):

- [1] 张友国. 积极稳妥推进碳达峰碳中和[J]. 中国产经, 2024(4): 28-31.  
ZHANG You-guo. Actively and steadily promoting carbon neutrality in peak carbon dioxide emissions[J]. Chinese Industry & Economy, 2024(4): 28-31.
- [2] 黄昕怡, 吴嘉仪, 林文浩, 等. 基于 GM(1, 1)模型的江苏省碳排放预测[J]. 黑龙江科学, 2022, 13(18): 26-28, 32.  
HUANG Xin-yi, WU Jia-yi, LIN Wen-hao, et al. Forecast of carbon emission of Jiangsu Province based on GM(1, 1) model[J]. Heilongjiang Science, 2022, 13(18): 26-28, 32.
- [3] 周树涛, 邱畅, 姚均业, 等. 基于 ARIMA 模型的碳排放预测及减排潜力[J]. 中国新技术新产品, 2023(2): 130-132.  
ZHOU Shu-tao, QIU Chang, YAO Jun-ye, et al. Carbon emission prediction and emission reduction potential based on ARIMA model [J]. New Technology & New Products of China, 2023(2): 130-132.
- [4] 赵雄飞, 李远利. 基于 LSTM 模型的中国 CO<sub>2</sub> 排放量预测影响因素分析[J]. 中国市场, 2021(22): 15-16.  
ZHAO Xiong-fei, LI Yuan-li. Analysis of influencing factors of CO<sub>2</sub> emission forecast in China based on LSTM model[J]. China Market, 2021(22): 15-16.
- [5] 陈欢, 牟瑛. “双碳”目标下重庆市碳排放影响因素及其预测研究[J]. 重庆工商大学学报(自然科学版), 2023, 40(2): 7-14.  
CHEN Huan, MU Ying. Research on the influencing factors and forecasting of carbon emissions in Chongqing under the “double carbon” target[J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2023, 40(2): 7-14.
- [6] 张学清, 李芳, 张绚, 等. 基于 CNN-LSTM 模型的中国碳排放量实时预测研究[J]. 中阿科技论坛(中英文), 2022(10): 71-75.  
ZHANG Xue-qing, LI Fang, ZHANG Xuan, et al. Research on real-time prediction of carbon emissions in China based on CNN-LSTM model[J]. China-Arab States Science and Technology Forum, 2022(10): 71-75.
- [7] 楼泽瑶. 基于 WTD-PPAR 的中国碳排放预测研究[J]. 软件导刊, 2024, 23(2): 99-105.  
LOU Ze-yao. Research on forecast of carbon emissions in China based on WTD-PPAR model[J]. Software Guide, 2024, 23(2): 99-105.
- [8] 王文佳, 潘昊, 王国刚. 基于 GWO-LSTM 模型的辽宁省工业碳排放预测及影响因素研究[J]. 环境科学与管理, 2024, 49(1): 28-33.  
WANG Wen-jia, PAN Hao, WANG Guo-gang. Prediction of industrial carbon peaking in Liaoning Province based on GWO-LSTM model[J]. Environmental Science and Management, 2024, 49(1): 28-33.
- [9] ABOYITUNGIYE J B, SURYANTO N A. Vulnerability of sub-Saharan Africa and Southeast Asian countries due to the carbon dioxide emissions: an assessment based on the

- STIRPAT model[J]. *International Journal of Environment and Pollution*, 2021, 69(1/2): 99.
- [10] SUTTHICHAIMETHEE P, CHATCHORFA A, SUYAPROM S. A forecasting model for economic growth and CO<sub>2</sub> emission based on industry 4.0 political policy under the government power: adapting a second-order autoregressive-SEM[J]. *Journal of Open Innovation: Technology, Market, and Complexity*, 2019, 5(3): 69.
- [11] MOORE T R, DAMON MATTHEWS H, CHAVAILLAZ Y. Linking historical and projected trends in extreme precipitation with cumulative carbon dioxide emissions [J]. *Atmosphere-Ocean*, 2024, 62(2): 165-182.
- [12] LIU D, CAO Y, YANG J, et al. SM-CycleGAN: crop image data enhancement method based on self-attention mechanism CycleGAN[J]. *Scientific Reports*, 2024, 14(1): 9277.
- [13] 李梅, 宁德军, 郭佳程. 基于注意力机制的 CNN-LSTM 模型及其应用[J]. *计算机工程与应用*, 2019, 55(13): 20-27.
- LI Mei, NING De-jun, GUO Jia-cheng. Attention mechanism-based CNN-LSTM model and its application [J]. *Computer Engineering and Applications*, 2019, 55(13): 20-27.
- [14] SAPUTRA A K, ERLANGGA E, TANJUNG T, et al. Review of deep learning using convolutional neural network model[J]. *Engineering Headway*, 2024, 70(82): 49-55.
- [15] QIAN F, TU Y, HOU C, et al. TLN-LSTM: an automatic modulation recognition classifier based on a two-layer nested structure of LSTM network for extremely long signal sequences[J]. *International Journal of Web Information Systems*, 2024, 20(3): 248-267.
- [16] 孔建国, 李亚彬, 张时雨, 等. 基于 CNN-LSTM-attention 模型航迹预测研究[J]. *航空计算技术*, 2023, 53(1): 1-5.
- KONG Jian-guo, LI Ya-bin, ZHANG Shi-yu, et al. Research on track prediction based on CNN-LSTM-attention model[J]. *Aeronautical Computing Technique*, 2023, 53(1): 1-5.
- [17] 张昱, 陈广书, 李继涛, 等. 基于 Attention 机制的 CNN-LSTM 时序预测方法研究与应用[J]. *内蒙古大学学报(自然科学版)*, 2022, 53(5): 516-521.
- ZHANG Yu, CHEN Guang-shu, LI Ji-tao, et al. Research and application of time series prediction method of CNN-LSTM based on attention mechanism[J]. *Journal of Inner Mongolia University (Natural Science Edition)*, 2022, 53(5): 516-521.
- [18] LIU C, PAN J, WANG J. An LSTM-based anomaly detection model for the deformation of concrete dams [J]. *Structural Health Monitoring*, 2024, 23(3): 1914-1925.
- [19] 唐清苇, 向月, 代佳琨, 等. 基于 CNN - LSTM 的风电场发电功率迁移预测方法[J]. *工程科学与技术*, 2024, 56(2): 91-99.
- TANG Qing-wei, XIANG Yue, DAI Jia-kun, et al. Wind farm power transfer forecasting method based on CNN-LSTM [J]. *Advanced Engineering Sciences*, 2024, 56(2): 91-99.
- [20] SELVANARAYANAN R, RAJENDRAN S, ALGBURI S, et al. Empowering coffee farming using counterfactual recommendation based RNN driven IoT integrated soil quality command system[J]. *Scientific Reports*, 2024, 14(1): 6269.
- [21] YI H, LIU J, XU W, et al. A graph neural network social recommendation algorithm integrating the multi-head attention mechanism[J]. *Electronics*, 2023, 12(6): 1477.
- [22] 杨璐霞, 王梦冉, 林兴亮, 等. 基于 CNN-LSTM 的工业出水水质预测模型[J]. *净水技术*, 2024, 43(5): 143-149.
- YANG Lu-xia, WANG Meng-ran, LIN Xing-liang, et al. Prediction model of industrial effluent quality based on CNN-LSTM[J]. *Water Purification Technology*, 2024, 43(5): 143-149.
- [23] 庞学丽, 宋坤, 姚红云, 等. 基于 VMD-ISSA-LSTM 的短时交通流预测研究[J]. *现代电子技术*, 2024, 47(8): 31-36.
- PANG Xue-li, SONG Kun, YAO Hong-yun, et al. Research on short-term traffic flow prediction based on VMD-ISSA-LSTM[J]. *Modern Electronics Technique*, 2024, 47(8): 31-36.
- [24] 魏健, 赵红涛, 刘敦楠, 等. 基于注意力机制的 CNN-LSTM 短期电力负荷预测方法[J]. *华北电力大学学报(自然科学版)*, 2021, 48(1): 42-47.
- WEI Jian, ZHAO Hong-tao, LIU Dun-nan, et al. Short-term power load forecasting method by attention-based CNN-LSTM[J]. *Journal of North China Electric Power University (Natural Science Edition)*, 2021, 48(1): 42-47.
- [25] 许栩. 基于 BP 神经网络模型的山东省碳排放量预测[J]. *中阿科技论坛(中英文)*, 2023(6): 21-27.
- XU Xu. Research on carbon emission forecasting in Shandong Province based on BP neural network model[J]. *China-Arab States Science and Technology Forum*, 2023(6): 21-27.

责任编辑:李翠薇