



中国测试
China Measurement & Test
ISSN 1674-5124, CN 51-1714/TB

《中国测试》网络首发论文

题目：基于注意力机制的CNN-ILSTM地铁站PM_{2.5}预测建模
作者：朱菊香，谷卫，罗丹悦，潘斐，张赵良
网络首发日期：2023-02-28
引用格式：朱菊香，谷卫，罗丹悦，潘斐，张赵良. 基于注意力机制的CNN-ILSTM地铁站PM_{2.5}预测建模[J/OL]. 中国测试.
<https://kns.cnki.net/kcms/detail/51.1714.TB.20230227.1553.002.html>



网络首发：在编辑部工作流程中，稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定，且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式（包括网络呈现版式）排版后的稿件，可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定；学术研究成果具有创新性、科学性和先进性，符合编辑部对刊文的录用要求，不存在学术不端行为及其他侵权行为；稿件内容应基本符合国家有关书刊编辑、出版的技术标准，正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性，录用定稿一经发布，不得修改论文题目、作者、机构名称和学术内容，只可基于编辑规范进行少量文字的修改。

出版确认：纸质期刊编辑部通过与《中国学术期刊（光盘版）》电子杂志社有限公司签约，在《中国学术期刊（网络版）》出版传播平台上创办与纸质期刊内容一致的网络版，以单篇或整期出版形式，在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊（网络版）》是国家新闻出版广电总局批准的网络连续型出版物（ISSN 2096-4188，CN 11-6037/Z），所以签约期刊的网络版上网络首发论文视为正式出版。

基于注意力机制的 CNN-ILSTM 地铁站 PM_{2.5} 预测建模

朱菊香¹, 谷卫², 罗丹悦², 潘斐², 张赵良¹

(1.无锡学院轨道交通, 江苏 无锡 214105; 2.南京信息工程大学自动化学院, 江苏 南京 210000)

摘要：为了提高 PM_{2.5} 的预测精度，提出了一种基于卷积神经网络 (CNN)、改进长短期记忆网络 (ILSTM) 和注意力机制 (Attention) 组合预测模型。ILSTM 删除了 LSTM 中的输出门，改进了其输入门和遗忘门，并引入了转换信息模块 (CIM) 以防止学习过程中的过饱和。该模型将一维卷积神经网络的特征提取和改进长短期记忆网络学习序列依赖性的能力相结合得到过去不同时间特征状态对未来 PM_{2.5} 浓度的影响，可以有效模拟 PM_{2.5} 在时间和空间上的依赖性，并通过注意力机制自动权衡过去的特征状态，进一步提升 PM_{2.5} 预测的准确度。实验结果表明：CNN-ILSTM-attention 模型的拟合度达到了 98.5%，与 LSTM 模型、CNN-LSTM 模型和 CNN-ILSTM 模型相比，分别提高了 26%、9.2%和 6.2%，具有较高的预测精度和应用价值。

关键词：卷积神经网络；改进长短期记忆网络；PM_{2.5} 浓度预测；注意力机制

中图分类号：TK223

文献标准码：A

Prediction modeling of PM_{2.5} in CNN-ILSTM subway station based on attention mechanism

ZHU Juxiang¹, GU Wei², LUO Danyue², PAN Fei², ZHANG Zhaoliang¹

(1. Rail Transit, Wuxi University, Wuxi 214105, China; 2. School of Automation, Nanjing University of Information Technology, Nanjing 210000, China)

Abstract: In order to improve the prediction accuracy of PM_{2.5}, a combined prediction model based on convolutional neural network (CNN), Improved Long short-term Memory Network (ILSTM) and Attention mechanism was proposed. ILSTM removes the output gate in LSTM, improves its input gate and forget gate, and introduces a transition information module (CIM) to prevent oversaturation during learning. One-dimensional convolutional neural network models of feature extraction and the ability to improve both short-term and long-term memory network learning sequence dependent state of different time of the past, the combination of characteristics of PM_{2.5} concentrations in the future, can effectively simulate the PM_{2.5} dependent on time and space, and through the attention mechanism automatic weighing the characteristics of the state in the past, To further improve the accuracy of PM_{2.5} prediction. The experimental results show that the fitting degree of CNN-ILSTM-Attention model reaches 98.5%, which is improved by 26%, 9.2% and 6.2%, respectively, compared with LSTM model, CNN-LSTM model and CNN-ILSTM model. It has high prediction accuracy and application value.

Key words: Convolutional Neural Network; Improved Long Short-Term Memory Network; PM_{2.5} Concentration Prediction; Attention Mechanism

基金项目：国家自然科学基金青年基金 (51206082); 江苏省自然科学基金 (be2015692); 江苏省高等学校自然科学基金项目 (21KJB460005)。

作者简介：朱菊香(1979-),女,江苏常州市人,副教授,硕士生导师,学历,研究方向为自动化及控制技术、检测技术。电话: 15251931986, E-mail: zjx@cwuxu.edu.cn。

通讯作者：张赵良(1979-),男,江苏无锡市人,硕士,高级工程师,研究方向为控制工程、仪器仪表。电话: 18262299609, E-mail: ZhangZL@cwuxu.edu.cn。

0 引言

近年来,随着我国城市轨道交通的快速发展,地铁因其方便、高效、安全等特点已成为日运载量最大的城轨交通工具^[1]。尽管地铁系统为人们的出行提供了方便,但是地铁系统中室内存在空气污染的问题,尤其是可吸入颗粒物(PM_{2.5})严重影响通勤者和地铁工作人员的身体健康^[2]。因此,精准预测 PM_{2.5} 浓度对于开发预警系统和调节通风系统以确保通勤者和地铁工作人员的健康具有重要的指导意义^[3]。

虽然目前地铁环境监控系统能够监测 PM_{2.5} 浓度的变化,但依然存在监控系统高成本、高功耗和低效率的问题^[4]。为了提高地铁站 PM_{2.5} 监测的有效性和准确性,降低监测成本,有必要建立更合理的预测模型。目前用于时间预测的模型有人工神经网络(ANN)、支持向量回归(SVR)、自回归综合移动平均(ARIMA)和随机森林回归(RFR)等。Nouri 等^[5]利用人工神经网络和主成分分析法来预测 PM_{2.5} 的浓度,虽然利用了 PM_{2.5} 的空间特征,但并未考虑 PM_{2.5} 在时间和空间上依赖性,导致预测精度较低;Li 等^[6]使用支持向量机回归(SVR)预测 PM_{2.5}。然而,由于 PM_{2.5} 数据的不稳定特性,SVR 很难达到高拟合度;刘拥民等^[7]使用 XGBoost-ARIMA 混合模型对城市中 PM_{2.5} 浓度进行预测,但无法处理环境数据中随机不确定性,因此预测精度较低;Feng 等^[8]结合随机森林回归(RFR)和 RNN 来分析和预测未来 24 小时的 PM_{2.5},但对于非线性 PM_{2.5} 的预测,RFR 存在过拟合的风险。由于地铁站室内空气环境的动态性质,递归神经网络(RNN)特别合适模拟 PM_{2.5} 的时间演变,因为 RNN 可以处理任意输入序列,从而保证学习时间序列的能力^[9,10]。然而,当传统 RNN 存在长时间滞后时,它可能会遇到“梯度消失”和“梯度爆炸”等问题,这些基于 RNN 的方法也没有充分利用空间特征^[11,12]。此外,不同时间特征形成的状态也会对未来的 PM_{2.5} 浓度产生不同的影响^[13]。LSTM 的出现可以缓解 RNN 训练过程中的“梯度消失”和“梯度爆炸”问题。白盛楠等人^[14]使用 LSTM 预测空气污染物的变化,由于预测模型单一,对特征值的提取不够充分,难以实现高精度预测;赵彦明等人^[15]提出基于时空相关性的 LSTM 的 PM_{2.5} 预测,但 LSTM 无法提取环境数据不同时间状态下的特征值,预测

精度较低。现有的研究没有考虑过去不同时间的特征状态对地铁站内污染物的影响,而只提取了历史数据的时间相关特征,导致预测精度较低。

为了解决上述问题,我们提出了一种基于注意力机制(Attention)的卷积神经网络(CNN)和改进长短期记忆(ILSTM)模型组合的 CNN-ILSTM-attention 混合神经网络模型,用于预测地地铁站室内 PM_{2.5} 的浓度。所提出的 CNN-ILSTM-attention 模型包括一维卷积神经网络(CNN)、改进长短期记忆(ILSTM)网络和注意力机制网络。为了克服传统 RNN 的“梯度消失”和“梯度爆炸”以及 LSTM 捕捉污染物相关时间序列数据的时空相关性和相互依赖性不充分的问题,对 LSTM 网络进行改进,删除了 LSTM 中的输出门,改进了 LSTM 中输入门和遗忘门,并引入了 CIM 来防止学习过程中的过饱和现象。同时,结合一维 CNN 目的是从地铁站室内空气中提取时空特征及局部空间相关性特征。注意力机制在预测模型中对过去的特征状态进行加权,以提高预测精度。因此,注意力机制应用于 CNN-ILSTM 模型,用于捕捉不同时间 PM_{2.5} 浓度的过去特征状态。通过无锡某地铁站采集的环境数据进行验证,实验表明所建立的 PM_{2.5} 预测模型可以有效地提高预测精度和泛化能力。

1 基于注意力机制的 CNN-ILSTM 的 PM_{2.5} 浓度预测

1.1 卷积神经网络(CNN)

CNN 是一种多层神经网络结构,具有强大的网格数据处理能力和深度监督学习结构,能够处理时间序列数据和图像数据。典型的 CNN 结构的主要包括输入层、卷积层、池化层、全连接层和输出层。输入层中的信息通过卷积层和池化层中的特征变换和提取进行处理,卷积层和池化层的本地信息由全连接层进一步集成,并通过输出层映射到输出信号^[16]。图 1 为一维 CNN 的网络结构,卷积层是 CNN 中最重要、最独特的一层,因为它可以通过卷积核提取输入变量的特征,与全连接层的神经网络不同,CNN 的卷积层只连接前一层的部分神经元,卷积核的规模小于输入矩阵的规模。卷积层使用卷积核运算而不是一般的矩阵运算来输出特征映射。变量特征中每个元素的计算公式如式(1)所示:

$$y_{i,j}^{out} = f_{cov}(\sum_{m=0}^k \sum_{n=0}^k w_{m,n} x_{i+m,j+n}^{in} + b) \quad (1)$$

式中， $y_{i,j}^{out}$ 是变量特征第*i*行和第*j*列中的输出值； $x_{i+m,j+n}^{in}$ 是输入矩阵第*i*行和第*j*列中的值； f_{cov} 是所选的激活功能； $w_{m,n}$ 是卷积核在*m*行和*n*列中的权重； b 是卷积核的偏

差。输入矩阵通常使用多个核进行卷积运算。每个卷积核将从输入矩阵中提取一个特征，并生成一个特征映射。之后，池化层减少了前一个特征映射的长度和宽度，并通过采样提高了计算效率。

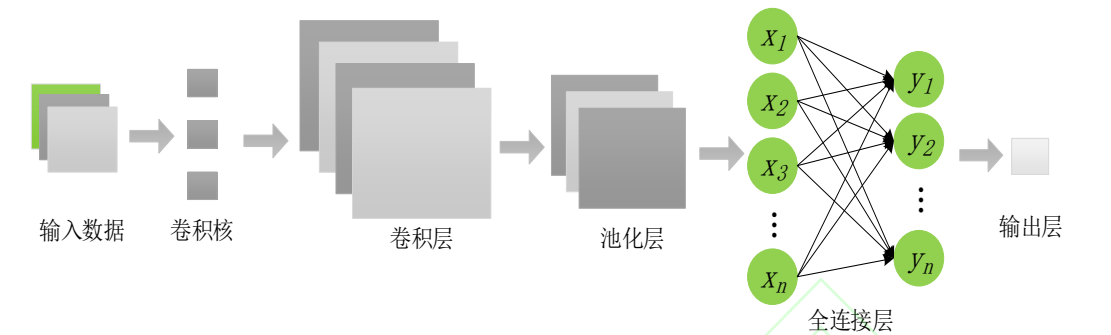


图 1 CNN 网络结构

1.2 改进长短期记忆网络（ILSTM）

长短期记忆神经网络（LSTM）是一种特殊的RNN^[17],LSTM模型具备了传统RNN的重复模块链式结构，区别于传统RNN的单一网络层，LSTM单元内部具有四个网络层。LSTM在挖掘序列数据的长期依赖关系方面具有很好的优势，将LSTM细胞加入神经元内部，即细胞中增加了门控结构，从而更新“细胞状态”。门控结构的作用是对流入的信息进行判断和甄别，来决定阻止或者允许信息的流入，从而选择性输入和丢弃信息，还能增加记忆单元，保存历史信息、神经元的长期状态。LSTM模型的内部结构如图2所示。门控递归单元（GRU）只有两个门。GRU将LSTM的输入门和遗忘门合二为一，称为更新门。基于门控技术，本文提出的ILSTM模型由输入门和遗忘门组成，如图3所示。与LSTM相比，ILSTM删除了输出门。与GRU相比，ILSTM结构更简单。LSTM、GRU和ILSTM的参数如表1所示。与LSTM相比，ILSTM将权重参数从8减少到4，将偏差参数从4减少到2。与GRU相比ILSTM将权重参数从6减少到4，将偏差参数从3减少到2。

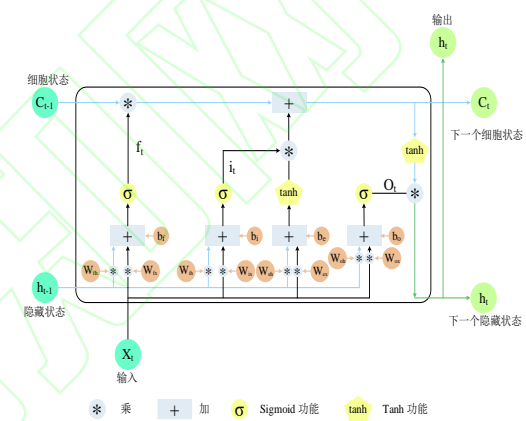


图 2 LSTM 神经元结构

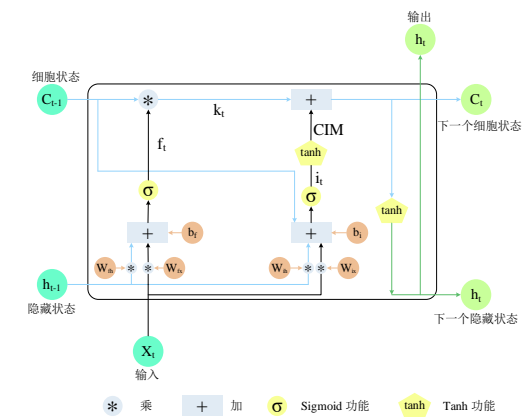


图 3 改进的 LSTM 神经元结构图

表 1 三种模型参数对比

LSTM	GRU	ILSTM
3 个门	2 个门	2 个门
8 重量参数	6 个重量参数	4 个重量参数
4 个偏置参数	3 个偏置参数	2 个偏置

在算法方面，ILSTM 将前一时刻的单元状态 c_{t-1} 加入到遗忘门算法中生成主线遗忘 k_t ，这将影响当前时间数据保留程度。此外，在更新当前时刻的细胞状态 c_t 时，引入了信息转换模块（CIM）以防止学习过程中的饱和现象。ILSTM 的遗忘门 f_t 是ILST 单元的主要组成部分，它可以控制哪

些信息应该被保留，哪些信息应该被遗忘。 $\sigma(x)$ 是一个 Sigmoid 函数，如公式（2）所示。 x_t 是第 t 个时间步长的输入数据。 h_{t-1} 是前一个时间步长 $t-1$ 的隐藏层。 W_{fh} 是遗忘门 h_{t-1} 的权重， W_{fx} 是 x_t 的权重。 b_f 是遗忘门的偏差，如公式（3）所示。

$$\sigma(x) = \frac{1}{1+e^{-x}} \quad (2)$$

$$f_t = \sigma(W_{fh} \cdot h_{t-1} + W_{fx} \cdot x_t + b_f) \quad (3)$$

主线遗忘 k_t 是由细胞状态 c_{t-1} 和 f_t 计算出来的。主线遗忘代表信息对当前细胞状态 c_t 的影响，其中 c_{t-1} 是从开始到前一刻的细胞状态信息，如公式（4）所示。

$$k_t = f_t \times c_{t-1} \quad (4)$$

输入门 i_t 控制当前输入数据 x_t 有多少流入存储单元，即有多少可以保存到 c_t 。与 LSTM 的输入门相比，ILSTM 在输入门算法中增加了 c_{t-1} ，即截止到前一刻的单元状态信息。 c_{t-1} 的引入使得模型的输入门对当前时间的数据保留具有记忆效应，如公式（5）所示， W_{ih} 和 W_{ix} 分别为输入门的 h_{t-1} 和 x_t 的权重， b_i 是输入门的偏置。

$$i_t = \sigma(W_{ih} \cdot h_{t-1} + W_{ix} \cdot x_t + c_{t-1} + b_i) \quad (5)$$

由于 Sigmoid 激活函数的特性，当 x 的值在 $[-3,3]$ 时，Sigmoid 激活函数的值将进入过饱和和区间。因此，在公式（5）中，当输入数据进入输入门的过饱和和区间时，其值不会发送明显变化，从而降低了学习的灵敏度。为了防止这种现象，在 ILSTM 模型中加入转换信息模块（CIM），如公式（6）所示。CIM 输出的值将大大降低过饱和度，显著差异使模型计算更容易识别，从而使模型学习更灵敏。

$$CIM = \tanh(i_t) \quad (6)$$

公式（7）表明， c_t 是从开始到现在保存的信息。

$$c_t = k_t + CIM \quad (7)$$

h_t 表示当前保存的信息。 c_t 控制通过 tanh 函数可以保存多少信息，如公式（8）所示。

$$h_t = \tanh(c_t) \quad (8)$$

ILSTM 推导公式如下：ILSTM 的提出是为了解决传统 RNN 的“梯度消失”和“梯度爆炸”以及 LSTM 捕捉污染物相关时间序列

数据的时空相关性和相互依赖性不充分的问题，以提高模型的预测精度。输入门和遗忘门使用两个参数矩阵 $[W_{fh}, W_{fx}]$ 和 $[W_{ih}, W_{ix}]$ 。记 $W_f = [W_{fh}, W_{fx}]$ ， $W_i = [W_{ih}, W_{ix}]$ 和 $W = [W_f, W_i]$ 。 W 的 L_t 函数是对应 h_t 的损失， L 表示总损失。至于 W 对 L 的导数，如公式（9）所示：

$$\frac{\partial L}{\partial W} = \sum_{t=1}^T \frac{\partial L_t}{\partial W} \quad (9)$$

RNN 通过公式（10）更新 W 参数：

$$W = W - \frac{\partial L_t}{\partial W} \quad (10)$$

其中 $\frac{\partial L_t}{\partial W}$ 可以写成公式（11）：

$$\frac{\partial L_k}{\partial W} = \frac{\partial L_k}{\partial h_k} \frac{\partial h_k}{\partial c_k} \frac{\partial c_k}{\partial c_{k-1}} \dots \frac{\partial c_2}{\partial c_1} \frac{\partial c_1}{\partial W} \quad (11)$$

公式（10）可以简化为公式（12）：

$$\frac{\partial L_k}{\partial W} = \frac{\partial L_k}{\partial h_k} \frac{\partial h_k}{\partial c_k} \dots \frac{\partial c_2}{\partial c_1} \frac{\partial c_1}{\partial W} = \frac{\partial L_k}{\partial h_k} \frac{\partial h_k}{\partial c_k} \left(\prod_{t=2}^k \frac{\partial c_t}{\partial c_{t-1}} \right) \frac{\partial c_1}{\partial W} \quad (12)$$

其中， c_t 如公式（13）所示：

$$c_t = f_t \times c_{t-1} + CIM \quad (13)$$

而 $CIM = \tanh(i_t)$ ，所以公式（14）可以写成公式（13）：

$$c_t = f_t \times c_{t-1} + \tanh(i_t) \quad (14)$$

而 c_t 的导数可以通过公式（15）获得：

$$\frac{\partial c_t}{\partial c_{t-1}} = f_t + (1 - \tanh(i_t)^2)(i_t)' \quad (15)$$

总损失可以写成公式（16）：

$$\frac{\partial L_k}{\partial W} = \frac{\partial L_k}{\partial h_k} \frac{\partial h_k}{\partial c_k} \left(\prod_{t=2}^k \left(f_t + (1 - \tanh(i_t)^2)(i_t)' \right) \right) \frac{\partial c_1}{\partial W} \quad (16)$$

此时记 x 等于公式（17）：

$$x = W_{fh} \cdot h_{t-1} + W_{fx} \cdot x_t + b_f \quad (17)$$

那么 f_t 可以写成公式（18）：

$$f_t = \sigma(W_{fh} \cdot h_{t-1} + W_{fx} \cdot x_t + b_f) = \sigma(x) \quad (18)$$

在记 y 等于公式（19）：

$$y = W_{ih} \cdot h_{t-1} + W_{ix} \cdot x_t + c_{t-1} + b_i \quad (19)$$

那么 i_t 可以写成公式（20）：

$$i_t = \sigma(W_{ih} \cdot h_{t-1} + W_{ix} \cdot x_t + c_{t-1} + b_i) = \sigma(y) \quad (20)$$

那么公式 (16) 可以写成公式 (21):

$$\frac{\partial L_k}{\partial W} = \frac{\partial L_k}{\partial h_k} \frac{\partial h_k}{\partial c_k} \left(\prod_{t=2}^k \left(\sigma(x) + \left(1 - \tanh(\sigma(y))^2 \right) (\sigma(y))' \right) \right) \frac{\partial c_l}{\partial W} \quad (21)$$

此时记 $z(x, y)$ 等于公式 (22):

$$z(x, y) = \sigma(x) + \left(1 - \tanh(\sigma(y))^2 \right) (\sigma(y))' \quad (22)$$

那么公式 (21) 可以写成公式 (23):

$$\frac{\partial L_k}{\partial W} = \frac{\partial L_k}{\partial h_k} \frac{\partial h_k}{\partial c_k} \left(\prod_{t=2}^k (z(x, y)) \right) \frac{\partial c_l}{\partial W} \quad (23)$$

此时函数的梯度如公式 (23) 所示。函数的梯度为 $\frac{\partial L_k}{\partial h_k} \frac{\partial h_k}{\partial c_k} \left(\prod_{t=2}^k (z(x, y)) \right) \frac{\partial c_l}{\partial W}$ ，当 $z(x, y)$ 大于 1 时，随着数据量的增加，梯度可能过大。当 $z(x, y)$ 太小时，梯度很容易消失。因此，ILSTM 模型可以在很大程度上缓解“梯度消失”和“梯度爆炸”的问题。

1.3 注意力机制

在堆叠的 ILSTM 层从特征中提取长期依赖关系后，其输出将被用作注意层的输入。每个隐藏状态的重要性由添加的注意层自动学习。注意机制可以简单地解释为加权总和。它首先计算输入特征的重要性；然后，应用 softmax 函数以确保所有权重之和为 1。权重和输入特征会相应地相乘，最终将它们相加得到输出。

注意机制可以自动为不同的时间阶段的不同特征分配不同权重，以区分它们的重要性，并且可以实现精确的时空可解释性。考虑过去状态的不同特征时的 $PM_{2.5}$ ，注意力机制允许模型捕捉 $PM_{2.5}$ 浓度最重要的部分。在提出的 CNN-ILSTM-attention 混合模型中，为了充分利用过去状态的信息，在 ILSTM 层中增加了一个注意层，它将过去不同特征状态的重要程度排序如下，其中 $H = [h_1, h_2, \dots, h_t]$ 是注意层中的特征状态

矩阵，公式表示为：

$$u_t = \tanh(W_h + b_t) \quad (24)$$

$$\alpha_t = \frac{\exp(u_t^T v)}{\sum_t \exp(u_t^T v)} \quad (25)$$

$$s = \sum_t \alpha_t h_t \quad (26)$$

式中， u_t 和 v 表示投影向量， α_t 是 h_t 的归一化权重， s 是注意层的加权输出向量。根据特征状态矩阵 H 中每个向量的重要性，等式 (24) 和 (25) 可以计算每个向量的归一化权重。等式 (26) 给出了加权向量 s 。这实现了在不同时间测量 $PM_{2.5}$ 特征状态的重要性。最终，加权向量 s 通过一层完全连接的网络来获得 $PM_{2.5}$ 浓度预测任务的专注度。

1.4 基于注意力机制 CNN-ILSTM 混合神经网络模型

CNN-ILSTM 结合了 CNN 和 ILSTM 的优点，并在此基础上加入注意力机制，提出 CNN-ILSTM-attention 混合神经网络模型，该模型包括一个输入层、一个卷积层、两个 ILSTM 层、一个注意层和一个全连接层。设置一维 CNN 层卷积核为 64，内核大小为 32，并在一维 CNN 中添加最大池化层，激活函数为 ReLU。卷积层和池化层中的 stride=1。设置双层 ILSTM 网络，每层的隐藏单元个数为 60，并在 ILSTM 网络中加入 Dropout，设置其值为 0.4，应用 Dropout 层可以防止模型过拟合。梯度下降的方式采用 Adam 训练算法，它能自动更新 ILSTM 网络的权重和偏置。本文提出的基于注意力机制的 CNN-ILSTM 模型具体实施步骤如下：首先，将数据输入模型，通过 CNN 的输入层、卷积层和池化层以提取输入数据的特征，将得到的特征展平成一维阵列，作为时间序列输入 ILSTM 层。其次，CNN 层的输出被送入第一个 ILSTM 层，该层捕获变量之间的长期依赖关系，并将每个 ILSTM 层的所有隐藏状态作为注意层的输入，注意层能够学习这些隐藏层状态的重要性。最后，通过全连接层得到 $PM_{2.5}$ 预测结果。图 4 为 CNN-ILSTM-attention 神经网络结构图。

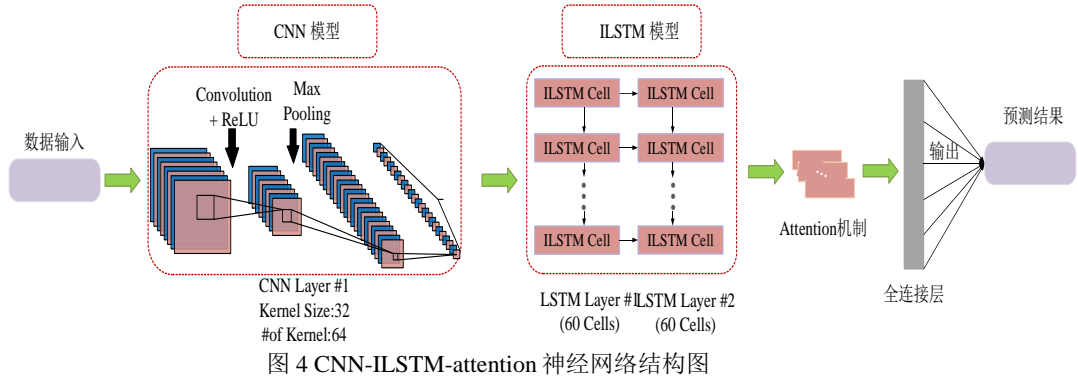


图 4 CNN-ILSTM-attention 神经网络结构图

2 基于注意力机制的 CNN-ILSTM 的 PM_{2.5} 预测建模

2.1 数据来源和处理

本文采用的数据是无锡某地铁站内的环境传感器采集的空气质量数据，包括 PM_{2.5}、PM₁₀、一氧化碳、二氧化硫、温度和湿度。环境传感器以 1min 为采集间隔，由传感器自动采集，共采集 1440 组数据。其中 1120 组数据作为训练样本，剩下 320 组数据作为测试样本检验 CNN-ILSTM-attention 模型对于 PM_{2.5} 浓度预测的准确性。

一旦将未经处理的原始数据输入模型，训练速度就会减慢，预测结果也会受到影响，因此为了进一步提高模型的预测精度和模型的收敛速度，所以需要输入量进行归一化处理。预测模型采用最大最小值标准化 (Min-Max Normalization)，将数据归一化到 [0,1] 之间，其转换公式 (27) 如下所示：

$$Y(x) = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (27)$$

式中： $\max(x)$ 为数据样本中的最大值； $\min(x)$ 为数据样本中的最小值； x 为数据样本中的任意值。

2.2 评估标准

为了评估 CNN-ILSTM-attention 模型的预测精度，选用均方根误差 (RMSE)、平均绝对误差 (MAE) 和拟合度 (r^2) 对算法模型进行评价。其计算公式如下：

$$\text{RMSE} = \sqrt{\frac{\sum_{t=1}^m (z_t - x_t)^2}{m}} \quad (28)$$

$$\text{MAE} = \frac{\sum_{t=1}^m |z_t - x_t|}{m} \quad (29)$$

$$r^2 = 1 - \frac{\sum_{t=1}^m (z_t - x_t)^2}{\sum_{t=1}^m (z_t - \bar{x}_t)^2} \quad (30)$$

式中， z_t 和 x_t 分别表示第 t 次测试的真实值和算法模型输出值， m 为测试总次数。

3 实验结果对比分析

3.1 不同时滞下的模型比较

一般的时间预测问题中，当面临长序列问题时，就会陷入瓶颈。一种解决方法就是通过引入注意力机制。注意力机制可以允许模型聚焦关注输入序列中需要的部分。引入了注意力机制后的模型，它可以将输入序列中相关部分的作用进行放大，以实现比没有注意力的模型更好的效果。通常，注意力机制的表现形式可以看作是一种加权求和。在本文中，为了提高集成神经网络模型的可扩展性和准确性，将注意力机制引入 CNN-ILSTM 网络模型中。

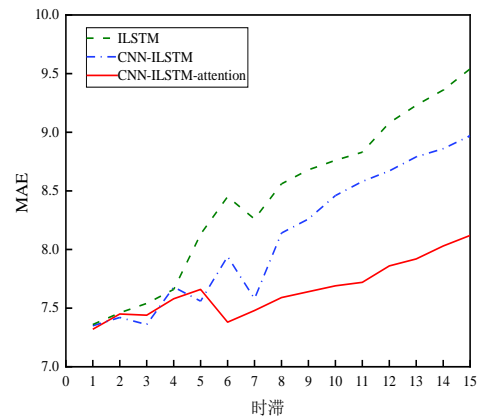


图 5 不同时滞下三种模型的 MAE 的比较

我们将 CNN-ILSTM-attention 模型、CNN-ILSTM 模型和 ILSTM 模型在 1 到 15 之间的时滞进行对比，以研究注意机制的有效性。时滞越大，提取有用特征的难度就越大。从图 5 可以看出，随着时滞的增加，MAE 变得越来越大。当时滞变大时，ILSTM 模型和 CNN-ILSTM 模型都迅速退

化，而 CNN-ILSTM-attention 模型性能保持稳定。结果表明，CNN-ILSTM-attention 模型具有较强的稳定性和鲁棒性，这是因为注意机制考虑了不同隐藏状态的重要性，并将更多注意力集中在更重要的状态上。

3.2 预测模型收敛性比较

设置合适的模型参数后，需要验证 CNN-ILSTM-attention 在训练过程中是否收敛。因此，计算了 CNN-ILSTM-attention 模型在 PM_{2.5} 浓度预测任务中的训练损失，并将结果与另外三种深度学习模型（LSTM、CNN-LSTM 和 CNN-ILSTM）进行了比较。所有预测模型的参数相同，使用数据归一化后的均方误差（MSE）作为损失函数进行训练。图 6 为四种模型训练损失函数，从图中可以看出，所有模型在 epoch=40 处收敛。从图中可以看出，在预测过程中 epoch=40 后，四个模型的 MSE 损失接近，但 CNN-ILSTM-attention 模型的 MSE 损失略小于 ILSTM、CNN-LSTM 和 CNN-ILSTM 三个模型且拥有更快的收敛速度。因此，CNN-ILSTM-attention 模型比 ILSTM 模型、CNN-LSTM 和 CNN-ILSTM 模型具有更好的收敛结果。

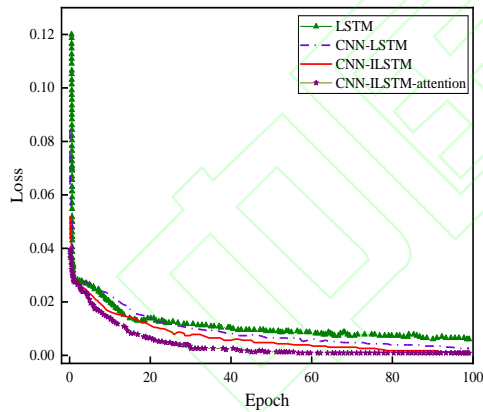


图 6 四种模型训练损失函数

3.3 预测模型性能比较

为了验证所提出的 CNN-ILSTM-attention 模型的有效性，本文两个方面进行对比。

a) 为了验证 CNN-ILSTM-attention 模型预测 PM_{2.5} 浓度的准确性，本文引入传统的回归模型（ARIMA、SVR 和 MLP）作为对比模型。实验结果如图 7、8、9、10 所示。从图中可以看出，传统的回归模型 ARIMA、SVR 和 MLP 的预测拟合度明显低于 CNN-

ILSTM-attention 模型。

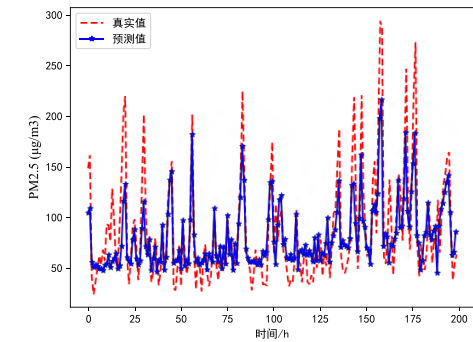


图 7 SVR 模型预测值和真实值对比

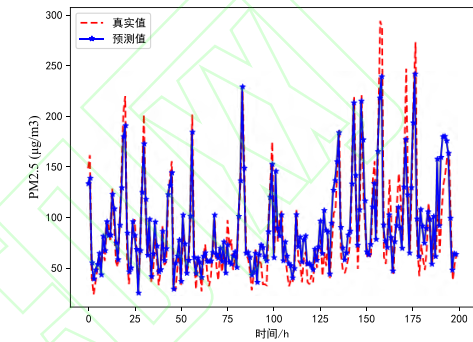


图 8 MLP 模型预测值和真实值对比

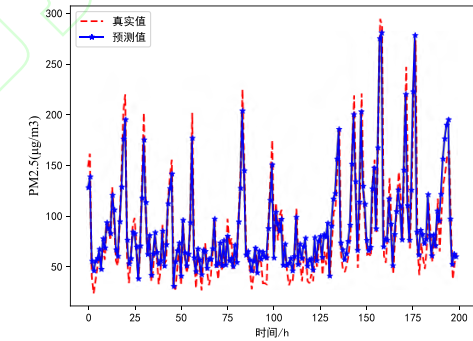


图 9 ARIMA 模型预测值和真实值对比

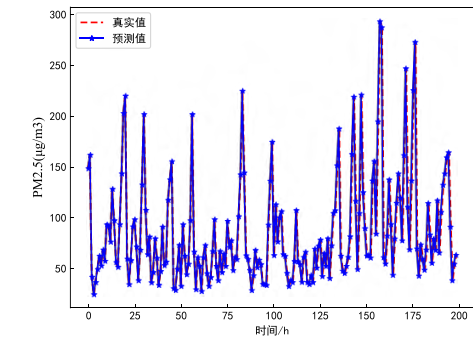


图 10 CNN-ILSTM-attention 模型预测值和真实值对比

表2为四种模型评价指标的对比。从表2可以看出, CNN-ILSTM-attention 预测模型的预测效果最好。与 SVR 模型相比, r^2 提高了 24.9%, RMSE 减少了 $15.04\mu\text{g}/\text{m}^3$, MAE 减少了 $11.1\mu\text{g}/\text{m}^3$ 。与 MLP 模型相比, r^2 增加了 14.2%, RMSE 减少了 $10.28\mu\text{g}/\text{m}^3$, MAE 减少了 $9.92\mu\text{g}/\text{m}^3$ 。与 ARIMA 模型相比, CNN-ILSTM-attention 模型的 r^2 提高了 12.1%, RMSE 减少了 $8.67\mu\text{g}/\text{m}^3$, MAE 减少了 $7.07\mu\text{g}/\text{m}^3$ 。因此, 相比较 ARIMA、SVR 和 MLP 等模型, CNN-ILSTM-attention 模型的对 $\text{PM}_{2.5}$ 浓度的预测具有较高的准确性。

表 2 四种预测模型评价指标对比

算法模型	RMSE($\mu\text{g}/\text{m}^3$)	MAE($\mu\text{g}/\text{m}^3$)	$r^2(\%)$
SVR	16.82	12.16	73.6
MLP	12.06	10.98	84.3
ARIMA	10.45	8.13	86.4
CNN-ILSTM-attention	1.78	1.06	98.5

b) 为了全面评估所提模型, 本文选用基于门控技术的递归神经网络模型 (LSTM) 以及混合预测模型 (CNN-LSTM、CNN-LSTM-attention、CNN-ILSTM) 与本文所提模型进行对比。

图 11 为 LSTM、CNN-LSTM、CNN-LSTM-attention、CNN-ILSTM 以及 CNN-ILSTM-attention 模型五种算法模型预测 $\text{PM}_{2.5}$ 浓度结果对比图, 图 12 为五种预测模型相对误差对比图。从图中可以发现 $\text{PM}_{2.5}$ 实际浓度值在不停的变化, LSTM 和 CNN-LSTM 两种预测模型能大致的捕捉 $\text{PM}_{2.5}$ 浓度变化规律, 依然存在预测误差较大的问题。CNN-LSTM-attention 模型、CNN-ILSTM 模型、CNN-ILSTM-attention 模型都能够准确捕捉 $\text{PM}_{2.5}$ 浓度变化规律, 但无法准确捕捉到当 $\text{PM}_{2.5}$ 浓度发生突然变化时的规律。在 CNN-ILSTM-attention 预测模型中, 注意层能够通过输入数据提取和存储在记忆单元中来改善深度学习模型中的记忆性能。从图中可以看出 CNN-ILSTM-attention 模型预测效果明显优于 LSTM、CNN-LSTM、CNN-LSTM-attention 和 CNN-ILSTM 这四种预测模型, 能够更好的捕捉到 $\text{PM}_{2.5}$ 浓度突然变化规律, 从而准确性的预测 $\text{PM}_{2.5}$ 的浓度。

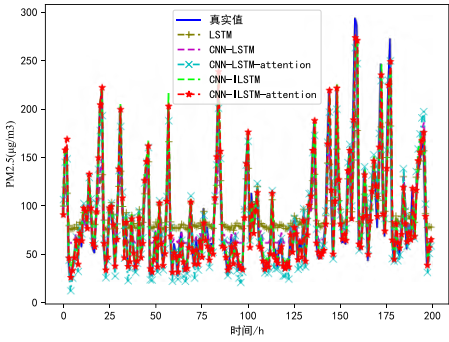


图 11 各模型预测结果对比

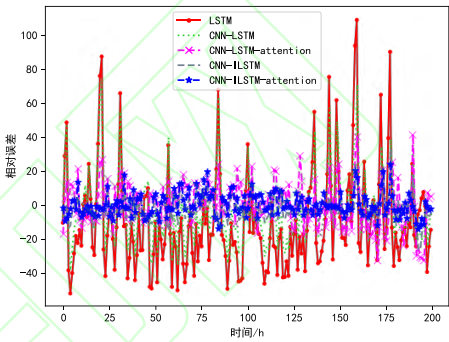


图 12 各模型相对误差对比

LSTM、CNN-LSTM、CNN-ILSTM 和 CNN-ILSTM-attention 等模型预测效果对比如表 3 所示。

表 3 不同模型预测效果评估

算法模型	RMSE($\mu\text{g}/\text{m}^3$)	MAE($\mu\text{g}/\text{m}^3$)	$r^2(\%)$
LSTM	17.05	12.89	72.5
CNN-LSTM	9.37	7.68	89.3
CNN-LSTM-attention	6.45	5.47	90.5
CNN-ILSTM	4.68	3.26	92.3
CNN-ILSTM-attention	1.78	1.06	98.5

从表 3 中可以看出, CNN-ILSTM-attention 模型拥有最高的拟合度以及最低的 RMSE 和 MAE。在 RMSE 和 MAE 指标上比 LSTM 模型和 CNN-ILSTM 模型降低了 $15.27\mu\text{g}/\text{m}^3$ 、 $2.9\mu\text{g}/\text{m}^3$ 和 $11.83\mu\text{g}/\text{m}^3$ 、 $2.2\mu\text{g}/\text{m}^3$ 。而 CNN-ILSTM-attention 模型在拟合度 r^2 为 98.5%, 较 LSTM 模型和 CNN-ILSTM 模型分别提高了 26%和 6.2%。因此可以得出: CNN-ILSTM-attention 模型预测

效果优于其他四种预测模型,呈现出更高的预测精度。

4 结束语

为了提高预测精度,提出了一种基于注意力机制的 CNN-ILSTM 模型来预测地铁站 $PM_{2.5}$ 的浓度。卷积层能够提取局部特征,而改进 LSTM 层能够获取长期依赖关系,注意力机制可以根据过去的时间捕捉不同特征状态的重要程度。实验结果表明,本文所建立的 CNN-ILSTM-attention 模型能够捕捉局部和全局的依赖关系,相比典型的 LSTM 模型、CNN-LSTM 模型以及 CNN-ILSTM 模型,具有更高的预测精度和泛化能力,这为精准预测 $PM_{2.5}$ 浓度提供理论和方法的支撑,这对于未来开发预警系统和调节通风系统以确保通勤者和地铁工作人员的健康具有重要的指导意义。

参考文献

- [1] GAO, XI, and WEIDE LI. "A graph-based LSTM model for $PM_{2.5}$ forecasting." *Atmospheric Pollution Research* 12.9 (2021): 101150.
- [2] 赵彦明.基于时空相关性的 LSTM 算法及 $PM_{2.5}$ 浓度预测应用[J].计算机应用与软件,2021,38(6):249-255+323.
- [3] 邢红涛,郭江龙,刘书安,等.基于 CNN-LSTM 混合神经网络模型的 NO_x 排放预测[J].电子测量技术,2022,45(2):98-103.
- [4] LOY-BENITEZ, JORGE, et al. "Sequential prediction of quantitative health risk assessment for the fine particulate matter in an underground facility using deep recurrent neural networks." *Ecotoxicology and environmental safety* 169 (2019): 316-324.
- [5] NOURI, AMIR, MEHDI GHANBARZADHE LAK, et al. "Prediction of $PM_{2.5}$ Concentrations Using Principal Component Analysis and Artificial Neural Network Techniques: A Case Study: Urmia, Iran." *Environmental Engineering Science* 38.2 (2021): 89-98.
- [6] LI, MIN, et al. "Prediction of $PM_{2.5}$ concentration based on the similarity in air quality monitoring network." *Building and Environment* 137 (2018): 11-17.
- [7] 刘拥民,罗皓懿,谢铁强.基于 XGBoost-ARIMA 方法的 $PM_{2.5}$ 质量浓度预测模型的研究及应用[J/OL].安全与环境学报:1-13[2022-07-03].
- [8] FENG, RUI, et al. "Recurrent Neural Network and random forest for analysis and accurate forecast of atmospheric pollutants: a case study in Hangzhou, China." *Journal of cleaner production* 231 (2019): 1005-1015.
- [9] 于伸庭,刘萍.基于长短期记忆网络-卷积神经网络(LSTM-CNN)的北京市 $PM_{2.5}$ 浓度预测[J].环境工程,2020,38(6):176-180+66.
- [10] ONG, BUN, KOJI ZETTSU, et al. "Dynamically pre-trained deep recurrent neural networks using environmental monitoring data for predicting $PM_{2.5}$." *Neural Computing and Applications* 27.6 (2016): 1553-1566.
- [11] 朱菊香,谷卫,罗丹悦,等.基于 PSO 优化 BP 神经网络的多传感器数据融合[J].中国测试,2022,48(8):94-100.
- [12] ALBIRMIZY, SHAHEEN, and BANAZ QADER. "Multivariate time series forecasting with lstm for madrid, spain pollution." *2019 International Conference on Computing and Information Science and Technology and Their Applications (ICCISTA)*. IEEE, 2019.
- [13] ABIRAMI, S., et al. "Hybrid spatio-temporal deep learning framework for particulate matter ($PM_{2.5}$) concentration forecasting." *2020 International conference on innovative trends in information technology (ICITIT)*. IEEE, 2020.
- [14] 白盛楠,申晓留.基于 LSTM 循环神经网络的 $PM_{2.5}$ 预测[J].计算机应用与软件,2019,36(1):67-70+104.
- [15] 高超,孙谊嫔,赵洪峰,等.基于 ICEEMDAN-ISSA-LSTM 的短期电力负荷预测[J/OL].中国测试:1-9[2022-07-05].
- [16] YAN, RUI, et al. "Multi-hour and multi-site air quality index forecasting in Beijing using CNN, LSTM, CNN-LSTM, and spatiotemporal clustering." *Expert Systems with Applications* 169 (2021): 114513.
- [17] XAYASOUK, GIVEOL LEE, et al. "Air pollution prediction using long short-term memory (LSTM) and deep autoencoder (DAE) models." *Sustainability* 12.6 (2020): 2570.