

韩飞龙, 冯淇萌, 周晓政, 等. 基于 CNN-SA-LSTM 的 PM2.5 浓度预测模型[J]. 智能计算机与应用, 2025, 15(7): 200-203.
DOI: 10.20169/j.issn.2095-2163.250730

基于 CNN-SA-LSTM 的 PM2.5 浓度预测模型

韩飞龙, 冯淇萌, 周晓政, 黄 远

(河北工程大学 信息与电气工程学院, 河北 邯郸 056038)

摘 要: PM2.5 是一种有害的环境污染物, 不仅会破坏自然环境, 还会危害人类健康。PM2.5 浓度的准确预报对控制大气污染、提高人们的生活质量具有重要意义。为了提高 PM2.5 浓度的预测精度, 本文提出了一种由卷积神经网络、自注意力机制、长短期记忆网络组合而成的深度学习模型, 其中卷积神经网络可以提取时序数据中隐含的深层特征、自注意力机制可以捕捉长期依赖、长短期记忆网络可以捕捉短期依赖, 从而提高 PM2.5 浓度的预测精度。实验分析表明本文提出的组合预测模型具有较高的预测精度。

关键词: 卷积神经网络; 自注意力机制; 长短期记忆网络; PM2.5 浓度预测

中图分类号: X513

文献标志码: A

文章编号: 2095-2163(2025)07-0200-04

PM2.5 concentration prediction model based on CNN-SA-LSTM

HAN Feilong, FENG Qimeng, ZHOU Xiaozheng, HUANG Yuan

(School of Information and Electrical Engineering, Hebei University of Engineering, Handan 056038, Hebei, China)

Abstract: PM2.5 is a harmful environmental pollutant that destroys the natural environment and harms human health. Accurate forecasting of PM2.5 concentration is of great significance to controlling air pollution and improving people's quality of life. In order to improve the prediction accuracy of PM2.5 concentration, this paper proposes a deep learning model composed of a convolutional neural network, a self-attention mechanism, and a long-short-term memory network. The convolutional neural network can extract the hidden meanings in time series data. The deep features and self-attention mechanism can capture long-term dependencies, and the long short-term memory network can capture short-term dependencies, thereby improving the prediction accuracy of PM2.5 concentration. Experimental analysis shows that the combined prediction model proposed in this article has high prediction accuracy.

Key words: convolutional neural network; self-attention mechanism; long short-term memory network; PM2.5 concentration prediction

0 引言

经济的持续增长、现代化的快速发展给人们的生活、工作和交通等方面带来了极大的便利,同时也造成了一定的环境污染问题^[1],引起了社会的广泛关注。PM2.5 是大气的主要污染物之一,不仅会污染周围空气环境,还会降低大气能见度,影响地球辐射收支,导致全球气候变化,甚至会严重威胁公众健康。相关研究表明,长期暴露于被细颗粒物污染的环境下,会增加呼吸道、心血管和肾脏疾病的风险。因此研发一套准确、可靠、稳定的 PM2.5 浓度预测系统对于有效进行大气生态环境监管和维护公众身体健康等方面具有重要意义。

近年来,随着人工智能和大数据的发展,深度学习技术受到了越来越多研究人员的关注,越来越多深度学习模型应用在时间序列预测领域,例如 Biancofiore 等^[2]就使用了循环神经网络 (Recurrent Neural Network, RNN)^[3]对 PM10 和 PM2.5 进行分析预测,实验结果充分证明了 RNN 在时间序列建模和分析中的优良性能,但 RNN^[3]在训练过程中可能难以捕捉序列中的长期依赖关系。由于 RNN 在训练过程中存在梯度消失或梯度爆炸的问题,潘永东等^[4]提出一种基于长短期记忆网络的 PM2.5 浓度预测方法,预测结果表明这种基于 LSTM 网络的 PM2.5 浓度预测方法,在一定程度上提高了对真实值的拟合效果。然而,单一的深度学习模型用于时间序列预测的

作者简介: 冯淇萌 (2000—), 女, 硕士研究生, 主要研究方向: 电力负荷预测; 周晓政 (1997—), 男, 硕士研究生, 主要研究方向: 目标检测; 黄远 (1987—), 男, 博士, 讲师, 主要研究方向: 数据挖掘, 预测算法研究。

通信作者: 韩飞龙 (2000—), 男, 硕士研究生, 主要研究方向: 深度学习, 时间序列预测。Email: 79293204@qq.com。

收稿日期: 2023-10-23

哈尔滨工业大学主办 ◆ 科技创新与应用

精度较低,无法全面有效地获取输入数据特征,因此,为了提高预测精度,学者们尝试将不同的深度学习模型进行组合,构建时间序列的组合预测模型,如熊思亦等^[5]将自适应噪声完备集合经验模态分解(CEEMDAN^[6])与LSTM相结合,提出一种海平面变化预测组合方法CEEMDAN-LSTM,相对于直接使用LSTM进行预测,这种方法具有最低的实验误差。郭宇辰等^[7]基于卷积神经网络(Convolutional Neural Network, CNN)与LSTM提出一种CNN-LSTM组合模型的碳价预测方法,将CNN-LSTM模型的预测结果与其他常用预测模型进行对比,结果表明CNN-LSTM模型在碳价预测中具有更高的预测准确性。同时,一种新的网络架构Transformer模型在自然语言处理^[8]、计算机视觉领域^[9]得到了广泛应用,Transformer模型全面依赖自注意力机制^[10]来捕捉序列中的依赖关系。自注意力机制可以直接关注输入序列中任何位置的信息,使其更容易捕获长距离依赖关系。

为了提高PM2.5浓度的预测精度,本文提出了一种基于卷积神经网络、自注意力机制和长短期记忆网络组合而成的短期PM2.5浓度预测模型,使用北京某监测站的空气质量数据来验证模型的性能,并与多种方法进行对比。实验结果表明,本文提出的组合预测模型具有较高的预测精度。

1 基本原理

1.1 卷积神经网络

卷积神经网络(CNN)是一种具有深度结构的前馈神经网络,自从LeCun等^[11]于1998年提出后,已被广泛应用于图像识别和计算机视觉领域。其独特优势在于能够自动且适应地从原始数据中学习空间层次上的特征,并通过堆叠多个处理层次来执行这一操作。这些层包括输入层、卷积层、池化层、全连接层和输出层。卷积神经网络的关键在于构建多个能提取数据特征的滤波器,核心是卷积层。卷积层依靠卷积核从输入数据中提取局部特征,并通过卷积核的滑动,获得输入数据的完整特征图。在经过多个卷积、池化操作后,卷积神经网络能挖掘数据中隐含的深层特征。

时序数据虽与图像数据在结构上有所不同,前者是按时间顺序组织的,而后者是在二维空间中组织的,但时序数据与图像数据中的空间模式相似,时间序列数据也包含有其内在的局部模式和结构。因此,CNN的卷积操作可以被应用于时间序列数据来

捕捉这些模式。

1.2 自注意力机制

基于自注意力机制的Transformer模型^[12]在自然语言处理领域大放异彩,这种自注意力机制可以用于识别和捕获一系列数据中的相似模式或特征,而且可以减少对外部信息的依赖,同时通过并行训练,则有助于时间序列预测模型对时间序列数据进行更有效的学习。自注意力机制将原始嵌入向量编码为3个新的向量,称为查询、键和值。然后,Self-Attention计算查询和关键字之间的相似度,并使用Softmax函数计算注意力得分。注意力的计算方法公式具体如下:

$$\text{Key} = X \times W_{\text{key}} \quad (1)$$

$$\text{Query} = X \times W_{\text{query}} \quad (2)$$

$$\text{Value} = X \times W_{\text{value}} \quad (3)$$

$$W_{\text{Attention}} = \text{Softmax} \left(\frac{\text{Query} \text{Key}^T}{\sqrt{d}} \right) \quad (4)$$

$$Y = W_{\text{Attention}} \times \text{Value} \quad (5)$$

其中, X 表示输入的数据; W_{key} 、 W_{query} 和 W_{value} 是权重矩阵;Softmax表示逻辑回归函数; d 表示 X 的维度。权重矩阵在训练过程中被学习,通过式(4)获得特征的注意力权重矩阵,最后通过式(5)得到注意力机制的输出。自注意力机制的计算过程如图1所示。

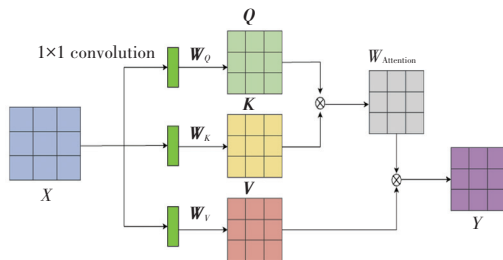


图1 自注意力机制的计算过程

Fig. 1 Calculation process of self-attention mechanism

1.3 长短期记忆网络

长短期记忆网络(LSTM)结构由Hochreiter等^[13]提出,其设计目的是为了解决循环神经网络(RNN)在学习长期依赖时遇到的问题。LSTM包含了特殊的门结构(例如遗忘门、输入门和输出门),使网络能够学习何时应该记住、更新或忘记其内部的状态。这种设计使得LSTM能够更为有效地对长序列进行处理。LSTM架构如图2所示。图2显示了LSTM的基本单元,称为cell, LSTM由许多连接的cell组成。LSTM的计算过程可描述为:

$$f_i = \sigma(W_f \cdot [h_{i-1}, x_i] + b_f) \quad (6)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{7}$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{8}$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \tag{9}$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{10}$$

$$h_t = \tanh(C_t) \odot o_t \tag{11}$$

接下来,对 LSTM 的整个过程展开分述如下:

首先,在更新新信息之前,LSTM 需要决定要从单元状态中遗忘哪些信息,这是通过遗忘门完成的。遗忘门也会查看当前的输入 x_t 和前一个时间步的隐藏状态 h_{t-1} ,输入到式(6)产生一个 0 到 1 之间的向量,这个向量表示每个单元状态位置应该保留多少旧信息。由此得到的较高的 f_t 表示保留更多的旧信息,而较低的值则表示遗忘更多。

然后,LSTM 决定要在单元状态中更新哪些信息。这是通过输入门来完成的。输入门会查看当前的输入 x_t 和前一个时间步的隐藏状态 h_{t-1} ,输入到式(7)得到 i_t ,这决定了每个单元状态应该更新多少。

接着,式(8)产生一个新的候选值向量 $\tilde{C}_t$,这个向量则提供了一个可能加入到状态中的新信息的选择。结合这 2 部分,就可以决定将哪些新信息加入到单元状态中。

此后,式(9)是更新后的新单元状态 C_t ,决定了哪些元素被遗忘,哪些元素将被保留。

最后,通过式(10)和式(11)计算出当前计算单元的隐藏状态。

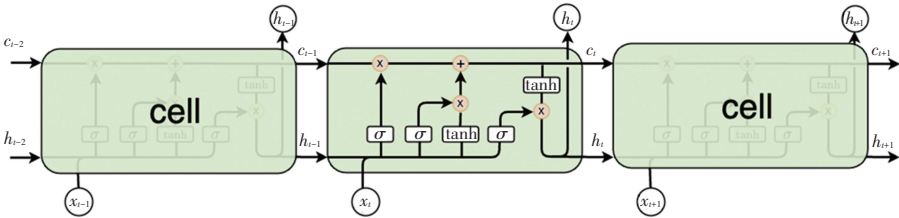


图 2 LSTM 计算示意图

Fig. 2 LSTM calculation diagram

2 基于 CNN-SA-LSTM 的组合预测模型

本文提出的 CNN-SA-LSTM 组合 PM2.5 浓度预测模型主要包括缺失值处理、模型训练、模型评估这 3 个步骤。首先,使用 Python 的第三方库对数据集缺失的时间进行补全,使用季节性分解函数对缺失的数据特征进行补全。然后,使用卷积神经网络能够有效地识别并提取空气质量数据中的局部模式和趋势,自注意力机制提取整个时间序列的全局依赖关系。最后,使用长短期记忆网络输出 PM2.5 浓度的预测值。整体流程图如图 3 所示。

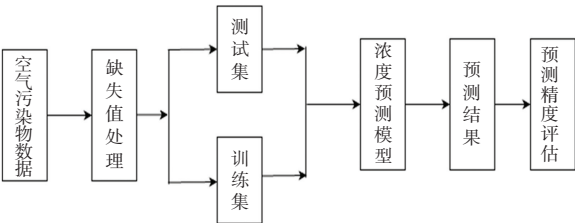


图 3 组合预测模型的流程示意图

Fig. 3 Flow chart of the combined prediction model

3 实验与分析

3.1 与其他模型进行比较

本文构建了 LSTM 模型、CNN 模型、CNN-LSTM

模型、Transformer 模型、SA 模型来分别预测 PM2.5 浓度。首先将处理缺失值后的数据进行窗口化处理,本文以前 5 h 预测未来 1 h 的 PM2.5 浓度,即把第 $(t-5)$ h、第 $(t-4)$ h、第 $(t-3)$ h、第 $(t-2)$ h、第 $(t-1)$ h 共 5 h 的 PM2.5 浓度数据和对应的气象数据代入模型进行训练。

3.2 评价方法

本文采用均方根误差 (RMSE)、平均绝对误差 (MAE)、平均绝对百分比误差 (MAPE) 和决定系数 (R^2) 来评价模型的预测效果。具体方法见表 1。

表 1 评价指标以及计算公式

Table 1 Evaluation indicators and calculation formulas			
指标	数学符号	计算方法	编号
均方根误差	RMSE	$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2}$	(1)
平均绝对误差	MAE	$MAE = \frac{1}{n} \sum_{i=1}^n \hat{y}_i - y_i $	(2)
平均绝对百分比误差	MAPE	$MAPE = \frac{1}{n} \sum_{i=1}^n \left \frac{\hat{y}_i - y_i}{y_i} \right $	(3)
决定系数	R^2	$R^2 = 1 - \frac{\sum_{i=0}^n (y_i - \hat{y}_i)^2}{\sum_{i=0}^n (y_i - \bar{y})^2}$	(4)

3.3 实验结果与分析

不同模型的 PM2.5 浓度预测结果如图 4 所示。各模型预测误差结果见表 2。从表 2 可以看出 LSTM 模型进行预测的 MAE 为 10.74, MAPE 为 0.87, RMSE 为 16.52, R² 为 0.78。在 LSTM、CNN、

Transformer 3 个单一模型中,LSTM 模型的预测性能最好。组合模型的各个评价指标均比单一模型的评价指标要好,本文提出的 CNN-SA-LSTM 组合模型性能也比 CNN-LSTM、SA-LSTM 预测性能都更好。

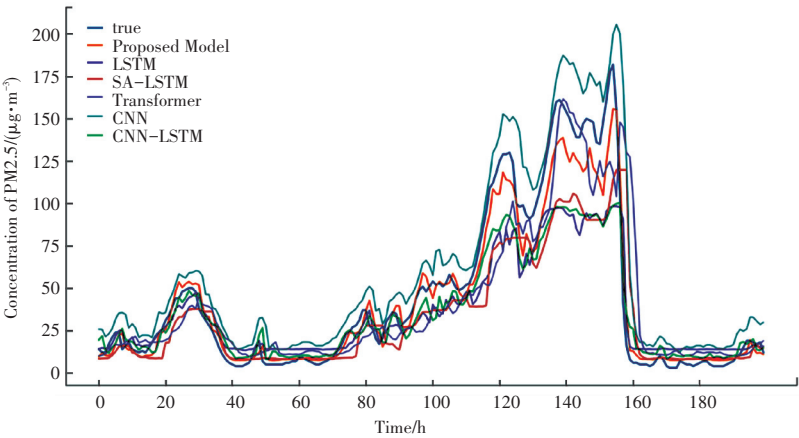


图 4 不同模型的 PM2.5 浓度预测结果

Fig. 4 PM2.5 concentration prediction results of different models

表 2 各预测模型误差结果

Table 2 Error results of each prediction model

模型	MAE	RMSE	MAPE	R ²
LSTM	10.74	16.52	0.87	0.78
CNN	14.80	18.63	1.13	0.73
CNN-LSTM	8.91	15.79	0.56	0.80
Transformer	12.36	19.89	0.81	0.69
SA-LSTM	10.98	16.95	0.56	0.77
CNN-SA-LSTM	5.75	9.08	0.39	0.93

4 结束语

本文提出了一种 CNN-SA-LSTM 组合预测模型,并采用北京某监测站的空气质量数据进行试验验证。该组合预测模型能够捕捉时间序列数据中长短期依赖的特点,也能对数据特征进行有效提取。实验结果表明,本文提出的 CNN-SA-LSTM 组合预测模型相比于单一的预测模型和 CNN-LSTM 和 SA-LSTM 模型具有较高的预测精度。

参考文献

[1] 张弦. 环境监测技术在生态环境保护中的运用研究[J]. 低碳世界,2022,12(12):22-24.
[2] BIANCOFIORE F, BUSILACCHIO M, VERDECCHIA M, et al. Recursive neural network model for analysis and forecast of PM10 and PM2. 5[J]. Atmospheric Pollution Research, 2017, 8(4): 652-659.

[3] MEDSKER L R, JAIN L C. Recurrent neural networks[J]. Design and Applications, 2001, 5(64-67): 2.
[4] 潘永东,曹骝,刘明. 基于 LSTM 网络的 PM2.5 浓度预测[J]. 金陵科技学院学报,2021,37(4):7-13.
[5] 熊思亦,熊永良. 基于 CEEMDAN-LSTM 组合方法的海平面变化预测分析[J]. 大地测量与地球动力学, 2023, 43(9): 899-903.
[6] TORRES M E, COLOMINAS M A, SCHLOTTHAUER G, et al. A complete ensemble empirical mode decomposition with adaptive noise[C]//Proceedings of the 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway,NJ: IEEE, 2011: 4144-4147.
[7] 郭宇辰,加鹤萍,余涛,等. 基于 CNN-LSTM 组合模型的碳价预测方法[J]. 科技管理研究,2023,43(11):200-206.
[8] 侯钰涛,阿布都克力木·阿布力孜,哈里旦木·阿布都克里木. 中文预训练模型研究进展[J]. 计算机科学, 2022, 49(7): 148-163.
[9] 李清格,杨小冈,卢瑞涛,等. 计算机视觉中的 Transformer 发展综述[J]. 小型微型计算机系统,2023,44(4):850-861.
[10] ZANG Haixiang, XU Ruiqi, CHENG Lilin, et al. Residential load forecasting based on LSTM fusing self-attention mechanism with pooling[J]. Energy, 2021, 229: 120682.
[11] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.
[12] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems(NIPS'17). New York: ACM, 2017: 6000-6010.
[13] HOCHREITER S, SCHMIDHUBER J. Long short-term memory [J]. Neural Computation, 1997, 9(8): 1735-1780.