

基于 PAM-BiLSTM 的 PM2.5 浓度预测模型

黄俊凯¹

HUANG Junkai

摘要

PM2.5 (细颗粒物) 是空气污染的主要成分之一, 长期暴露于高浓度的 PM2.5 环境中, 对人类健康和环境都有严重危害, 包括呼吸系统疾病、心血管疾病和早逝等。因此, 基于时间序列数据准确预测 PM2.5 浓度对于提前决策和风险管理至关重要。提出一种基于金字塔注意力机制 (PAM) 和双向长短期记忆网络 (BiLSTM) 的 PM2.5 浓度预测模型。实验结果表明, PAM-BiLSTM 模型在预测精度和效率方面具有显著优势, 同时消融实验验证了 PAM 模块的有效性, 在多个评价指标上优于 BiLSTM 模型, 对于环境监测和公共健康保护具有重要意义。

关键词

PM2.5; 时间序列数据; 金字塔注意力机制; 双向长短期记忆网络

doi: 10.3969/j.issn.1672-9528.2024.09.014

0 引言

PM2.5 (细颗粒物) 是空气污染的主要成分之一, 其直径小于 2.5 微米, 能够长时间悬浮在空气中。由于粒径微小, PM2.5 能深入肺部, 甚至进入血液, 对人体健康造成严重威胁^[1]。准确预测 PM2.5 浓度对环境监测和公共卫生保护至关重要。

目前, PM2.5 浓度预测的研究工作以深度学习方法为主。Chen 等人^[2]提出了一种用于短期 PM2.5 预测的深度学习模型, 该模型的显著特征是模型架构中的卷积是因果的, 其中时间步的输出仅与前一层相同或更早时间步的分量进行卷积。Wu 等人^[3]提出了一种加权神经网络 (STWNN), 以应对站点稀疏且不均匀的地区长期 PM2.5 预测不佳的挑战。Zhang 等人^[4]提出了一种基于时空图注意力循环神经网络和灰狼优化的新方法 (GWO-GART) 来针对城市空间特征的提取。Wang 等人^[5]针对地铁系统中的 PM2.5 监测系统, 提出了一种集成 KPCA-AdaBoost-LSTM 模型。Wang 等人^[6]结合图形卷积模块、时间卷积模块、线性模块开发了一种新型的 PM2.5 时空预报框架, 它能够预测未来多个站点和多个时间步长的 PM2.5 浓度。杨国亮等人^[7]提出基于 PSO-VMD-LSTM 的组合模型, 利用以包络熵为适应度函数的 PSO 机器学习算法对 VMD 分解过程中的惩罚因子和模态个数进行迭代寻优, 基于分解处理后的数据进行 LSTM 深度学习模型预测。

然而, 现有方法在捕捉多尺度时间依赖关系和提高预测精度方面仍存在不足, 影响了实际应用的效果。同时, 传统

LSTM 模型是单向的, 只考虑前向时间依赖关系, 忽略了未来信息, 也容易出现过拟合问题。因此, 研究一种能够高效捕捉多尺度时间依赖关系的预测模型具有重要意义。

为了解决上述问题, 本文提出了一种全新的模型架构, 结合金字塔注意力机制 (pyramidal attention module, PAM) 与双向长短期记忆网络 (BiLSTM) 的 PM2.5 浓度预测模型。金字塔注意力模块 (PAM) 通过多分辨率特征提取和跨尺度注意力机制, 有效捕捉时间序列数据中的多尺度特征。双向长短期记忆网络通过前向和后向两个方向的信息传递, 进一步提升了提取时间依赖关系的能力和对复杂时序数据的理解能力。通过结合 PAM 和 BiLSTM 的优势构建一个高效、准确的模型。

1 金字塔注意力机制

金字塔注意力机制是 Pyraformer 的核心组件, 它通过多分辨率方式描述时间序列的时间依赖关系, 极大地提升了模型的效率和效果^[8]。它用于在金字塔图中的节点之间进行信息交换, 以捕捉不同范围的时间依赖关系。PAM 的输入是一个金字塔图, 其中每个节点代表一个时间步长的特征, 而节点之间的连接表示它们之间的关系。通过这种金字塔结构, PAM 可以在不同的时间尺度上进行信息交换, 从而提高时间序列建模和预测的准确性。以下是 PAM 的工作原理。

1.1 多头注意力机制

使用多头注意力机制, 通过不同的注意力头 (head) 对输入特征进行并行处理, 以捕捉不同子空间的特征表示。PAM 的注意力机制可以表示为:

1. 南宁师范大学物理与电子学院 广西南宁 530100

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

式中： Q 、 K 和 V 分别为查询（query）、键（key）和值（value）的特征表示。通过计算查询和键之间的相似度，然后将其应用于值，可以得到每个查询的加权值。这样，每个节点可以根据与其他节点的相似度来获取上下文信息。

1.2 输入特征的线性变换

对输入特征进行线性变换，线性变换的公式如式

(2) ~ (4)，得到查询 Q 、键 K 和值 V ：

$$Q = W_q X \quad (2)$$

$$K = W_k X \quad (3)$$

$$V = W_v X \quad (4)$$

式中： W_q 、 W_k 和 W_v 是可训练的权重矩阵， X 是输入特征。

1.3 分头处理

将线性变换后的查询、键和值按头数进行拆分，以便并行处理。

1.4 缩放点积注意力

计算 Q 和 K 的点积，并进行缩放，得到注意力权重，再对注意力权重应用 Softmax 函数，得到归一化的注意力分数，注意力权重和归一化注意力分数的公式为：

$$\text{scaledattentionlogits} = \frac{Q_{split}K_{split}^T}{\sqrt{d_k}} \quad (5)$$

$$\text{attentionweights} = \text{softmax}(\text{scaledattentionlogits}) \quad (6)$$

1.5 加权求和值

根据注意力分数对值进行加权求和，公式如 (7) 所示，得到每个查询的上下文信息：

$$\text{scaledattention} = \text{attentionweights} \cdot V_{\text{split}} \quad (7)$$

1.6 拼接与线性变换

将多头注意力的结果拼接起来，并进行线性变换，公式如 (8) 所示，得到最终的输出：

$$\text{output} = W_o \cdot \text{reshape}(\text{transpose}(\text{scaledattention})) \quad (8)$$

式中： W_o 是输出的权重矩阵。

2 PAM-BiLSTM 模型设计

2.1 PAM-BiLSTM 模型构建

模型的整体架构由 PAM、BiLSTM 和一个全连接层组成，可以捕捉 PM2.5 浓度预测中的长短期时间依赖关系及多尺度特征。BiLSTM 层作为主要的序列建模工具，利用其双向架构有效地捕捉了数据中的长期和短期时间依赖关系，从而在预测过程中充分考虑了历史数据的影响^[9]。PAM 层则通过多头注意力机制，增强了模型对不同时间尺度特征的感知能力，使其能够更精确地识别和利用数据中的重要信息。这种组合

使模型能够在面对复杂的 PM2.5 浓度变化时，更加准确地进行预测和决策支持。PAM-BiLSTM 的模型结构图如图 1 所示。

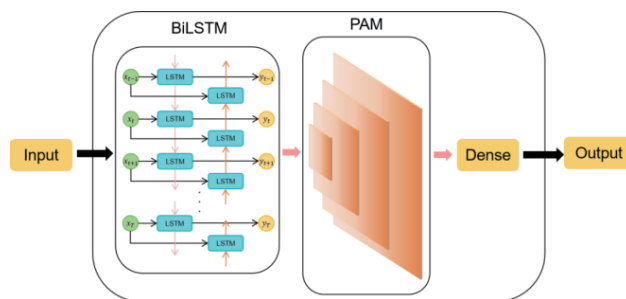


图 1 PAM-BiLSTM 模型结构图

2.2 数据集预处理

本文的数据来自 Liang 等人提供的数据集，使用了位于北京（东经 116.47°、北纬 39.95°）的美国大使馆每小时的 PM2.5 读数，以及来自 weather.ncrew.org 的北京首都国际机场（BCIA）每小时的气象测量数据^[10]。数据时间段为 2010 年 1 月 1 日至 2014 年 12 月 31 日。数据集信息如表 1 所示。

表 1 数据集信息介绍

数据集	Beijing PM2.5.UCI Machine Learning Repository
实例	43824
特征	11
时间段	2010 年 1 月 1 日至 2014 年 12 月 31 日
PM2.5	PM2.5 浓度 (μg/m ³)
DEWP	露点 (°C)
TEWP	温度 (°C)
PRES	压力 (hPa)
cbwd	组合风向
lws	累计风速 (m/s)
ls	累计降雪小时数
lr	累计降雨小时数

为了提高模型的预测性能，必须对原始数据集进行预处理，首先使用贝叶斯网络去除数据中的缺失值和异常值，确保输入数据的质量^[11]。再对数据集进行特征提取，提取与 PM2.5 相关的特征，最后对所有输入特征进行归一化处理，使取值范围在 0 到 1 之间，消除不同特征之间的量纲差异，提高模型训练的稳定性 and 收敛速度。其中，将数据归一化的算法为：

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (9)$$

式中： x 是原始数据； x' 是归一化后的数据； x_{\max} 是数据集的最大值； x_{\min} 是数据集的最小值。

2.3 模型预测步骤

使用 PAM-BiLSTM 模型对每小时的 PM2.5 浓度的预测

步骤简述如下。

(1) 数据预处理：首先，对原始的数据集进行预处理，包括清洗数据、特征选择、数据标准化。

(2) 数据集分割：将预处理好的数据集划分为训练集和测试集，数据集的时间跨度为 5 年，本文将数据集的前 4.5 年划分为训练集，后 0.5 年划分为测试集。

(3) 构建模型：使用双向长短期记忆网络和金字塔注意力机制构建预测模型，并对其进行训练。**BiLSTM 层**：使用 100 个隐藏单元的双向 LSTM 层，捕捉序列数据的长短期依赖关系。**全连接层**：添加一个全连接层，将 LSTM 层的输出进行线性变换。**PAM 层**：进一步提取不同时间尺度的特征，提高模型的预测精度。**全连接层**：输出 PM2.5 浓度的预测值。

(4) 编译并训练模型：设定优化器和损失函数，使用训练集进行模型训练。

(5) 模型预测：在模型训练完成后，对模型进行评估，并进行 PM2.5 浓度预测。

(6) 结果分析：将预测值与实际值进行比较，计算各类评价指标。

3 实验与结果分析

3.1 实验环境

本文中实验采用的计算机处理器为 Intel Core i7-12700H，内存为 16 GB，操作系统的版本为 Windows11（64 位），实验使用的开发编辑器为 PyCharm 2024.1，实验使用的 Python 版本为 3.7.1。实验中 PAM-BiLSTM 模型的参数设置为：使用双向 LSTM 设置 100 个隐藏单元，return_sequence 设置为 True，返回每个时间步的输出，输入数据的时间步数为 1，每个时间的特征步数为 8。输入金字塔层的参数中，注意力头的数量设置为 4，模型的维度为 8。学习率优化算法选择“Adam”，学习率设置为 0.000 1，迭代次数设置为 300。

3.2 评价指标

在时间序列预测任务中，选择合适的评价指标是衡量模型性能的重要步骤。本文选择均方根误差（root mean squared error, RMSE）、平均绝对误差（mean absolute error, MAE）、平均绝对百分比误差（mean absolute percentage error, MAPE）和决定系数（coefficient of determination, R^2 ）作为评价本次实验的指标。四个评价指标的计算方式为：

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (10)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (11)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (12)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (13)$$

式中： y_i 是真实值， \hat{y}_i 是预测值， \bar{y}_i 为平均值， n 是样本数量。RMSE 对于大的误差更加敏感，因此能够放大较大的误差；MAE 直接反映了预测值与真实值之间的平均偏差；MAPE 提供了一个相对误差的度量，可以用于不同量级数据的对比。决定系数 R^2 的值越接近 1，表示模型对数据的解释程度越高；值越接近 0，表示模型对数据的解释程度越低。

3.3 结果分析

使用测试集对训练完成的 PAM-BiLSTM 模型进行模型评估，选取测试集前 500 个小时的预测数据进行评估，得到的预测值与实际值的对比图如图 2 所示，从图中可以看出，该模型的预测结果较为接近实际数据，预测误差较小，对数据的解释能力强，具有较高的预测准确度和稳定性。

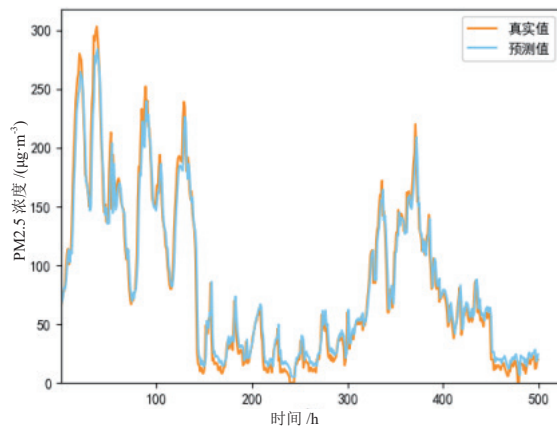


图 2 PAM-BiLSTM 预测结果

通过对比不同方法的预测效果，可以客观评估模型在特定浓度预测任务中的表现。本文将 PAM-BiLSTM 模型与 LSTM、CNN-LSTM 和 CNN-BiLSTM 模型进行比较。对比的模型将在相同的实验条件下进行，选取 RMSE、MAE、MAPE、 R^2 作为评价指标。比较的结果如表 2 所示。

表 2 不同模型的预测性能比较

Model	RMSE	MAE	MAPE	R^2
PAM-BiLSTM	22.375	12.249	0.221	0.969
CNN-BiLSTM	24.056	13.622	0.259	0.962
CNN-LSTM	24.296	13.822	0.269	0.962
LSTM	27.183	14.512	0.251	0.957

3.4 消融实验

消融实验（ablation study）是评估模型中各个组件的重要方法，通过逐步去除或替换模型中的特定组件，分析这些组件对模型性能的影响，从而验证各部分在整个模型中的贡

献。以下是针对 PAM-BiLSTM 模型的消融实验设计。构建两个模型，一个包含 PAM (PAM-BiLSTM 模型)，另一个不包含 PAM (仅 BiLSTM 模型)。对比两个模型在 RMSE、MAE、MAPE、相关系数和 R^2 评价指标上的表现，分析 PAM 的引入对模型性能的提升。实验结果如表 3 所示。PAM 在捕捉多尺度时间依赖关系方面有效，特别是在长序列预测任务中，PAM-BiLSTM 模型应在各项评价指标上表现优于 BiLSTM 模型。

表 3 PAM-BiLSTM 模型与 BiLSTM 模型的预测性能比较

Model	RMSE	MAE	MAPE	R^2
PAM-BiLSTM	22.375	12.249	0.221	0.969
LSTM	23.535	13.217	0.234	0.965

4 总结

本文提出了一种结合金字塔注意力机制和双向长短期记忆网络的 PM2.5 浓度预测模型，以应对现有方法在捕捉多尺度时间依赖关系和提高预测精度方面的不足。通过引入 PAM，有效地捕捉了时间序列数据中的多尺度特征，从而提升了模型的预测能力。同时，BiLSTM 进一步加强了对复杂时间依赖关系的提取，确保了模型在处理长时间序列时的性能。

在实验部分，本文详细介绍了数据集的预处理方法、模型的构建步骤以及使用的评价指标。通过与传统的 LSTM、CNN-LSTM 和 CNN-BiLSTM 模型进行对比，实验结果表明，PAM-BiLSTM 模型在 RMSE、MAE、MAPE 和 R^2 等多个评价指标上表现优异，预测误差显著降低，模型的解释能力和稳定性得到提升。此外，本文通过消融实验验证了 PAM 在模型中的重要性。BiLSTM 模型在各项评价指标上均低于包含 PAM 的模型，进一步证明了 PAM 在捕捉多尺度时间依赖关系和提升模型性能方面的有效性。

总的来说，本文提出的 PAM-BiLSTM 模型在 PM2.5 浓度预测任务中表现出色，为环境监测和公共卫生保护提供了一种高效、准确的工具。未来的研究可以在以下几个方面进一步探索：首先，继续优化 PAM 的结构以进一步提升模型的计算效率；其次，结合更多种类的环境数据，提高模型的泛化能力；最后，采用可解释性技术分析模型决策过程，增强模型的透明度和可信度。

参考文献:

[1]PING L, WANG Y, LU Y, et al.Tracing the sources of PM2.5-related health burden in China[J].Environmental pollution, 2023, 327:121544.

[2]CHEN Y, HUANG L, XIE X, et al. Improved prediction of hourly PM2.5 concentrations with a long short-term memory and spatio-temporal causal convolutional network deep learning model[J].Science of the total environment, 2024, 912: 168672.

[3]WU Y, XU Z, XU L, et al.An improved deep learning approach considering spatiotemporal heterogeneity for PM2.5 prediction: a case study of Xinjiang, China[J].Atmosphere, 2024, 15(4): 460.

[4]ZHANG C, WANG S, WU Y, et al.A long-term prediction method for PM2.5 concentration based on spatiotemporal graph attention recurrent neural network and grey wolf optimization algorithm[J].Journal of environmental chemical engineering, 2024, 12(1):111716.

[5]WANG J, WANG D, ZHANG F, et al.Soft sensor for predicting indoor PM2.5 concentration in subway with adaptive boosting deep learning model[J].Journal of hazardous materials, 2024, 465: 133074.

[6]WANG J, WU T, MAO J, et al.A forecasting framework on fusion of spatiotemporal features for multi-station PM2.5[J].Expert systems with applications, 2024,238:121951.

[7]杨国亮, 余华声, 黄聪. 基于优化组合模型的 PM2.5 浓度预测 [J]. 计算机工程与设计, 2023,44(10):3132-3137.

[8]LIU S Z, YU H, LIAO C, et al. Pyraformer: low-complexity pyramidal attention for long-range time series modeling and forecasting[C//OL]//ICLR 2022 Oral.[2024-05-21].https://openreview.net/forum?id=0EXmFzUn5I.

[9]SIAMI-NAMINI S, TAVAKOLI N, NAMIN A S.The performance of LSTM and BiLSTM in forecasting time series[C//2019 IEEE International Conference on Big Data (Big Data). Los Angeles, CA, USA: IEEE, 2019:3285-3292.

[10]LIANG X, ZOU T, GUO B, et al. Assessing Beijing's PM 2.5 pollution: severity, weather impact, APEC and winter heating[J].Proceedings of the royal society A: mathematical, physical and engineering sciences, 2015,471(2182):20150257.

[11]许志城, 徐康康, 杨海东, 等. 一种基于朴素贝叶斯的大段不等长缺失数据填补方法: CN110826718A[P].2020.

【作者简介】

黄俊凯 (1999—), 男, 广西柳州人, 硕士研究生, 研究方向: 深度学习、时间序列预测。

(收稿日期: 2024-06-17)