

基于特征提取和LSTM的PM_{2.5}浓度预测模型

舒莹¹, 胡宸滔¹, 铁冶欣^{1,2*}, 丁成富³

¹浙江理工大学信息科学与工程学院, 浙江 杭州

²浙江理工大学科技艺术学院, 浙江 绍兴

³聚光科技(杭州)股份有限公司, 浙江 杭州

收稿日期: 2023年3月20日; 录用日期: 2023年5月18日; 发布日期: 2023年5月25日

摘要

PM_{2.5}是空气污染中对人体危害最大的一类污染物之一, 对PM_{2.5}的准确预测可以为人们社会活动的决策制定提供可靠依据。使用自编码器(Auto-encoder)能够达到较高的预测精度, 但是需要训练的参数量大, 所需的计算资源也会增加, 于是本文提出了一种待训练参数比Auto-encoder减少了24%的PM_{2.5}预测模型FE-LSTM (Feature Extraction-LSTM)。FE-LSTM模型基于SE注意力机制、卷积神经网络(CNN)和长短期记忆网络(LSTM)等模块, 先使用CNN提取出输入张量的初始特征, 再通过SE注意力对特征张量按通道加权, 通过全连接和重构后使用LSTM得出污染浓度的预测值。在北京PM_{2.5}数据集上, 对FE-LSTM模型及对比模型进行了训练和测试, 结果表明, 本文所提出的FE-LSTM模型的预测精度优于其他对比模型。

关键词

SE注意力, 卷积神经网络, 长短期记忆网络, PM_{2.5}预测

PM_{2.5} Concentration Prediction Model Based on Feature Extraction and LSTM

Ying Shu¹, Chentao Hu¹, Zhixin Tie^{1,2*}, Chengfu Ding³

¹School of Information Science and Technology, Zhejiang Sci-Tech University, Hangzhou Zhejiang

²School of Science and Technology, Zhejiang Sci-Tech University, Shaoxing Zhejiang

³Juguang Technology (Hangzhou) Co., Ltd., Hangzhou Zhejiang

Received: Mar. 20th, 2023; accepted: May 18th, 2023; published: May 25th, 2023

*通讯作者。

文章引用: 舒莹, 胡宸滔, 铁冶欣, 丁成富. 基于特征提取和 LSTM 的 PM_{2.5} 浓度预测模型[J]. 建模与仿真, 2023, 12(3): 2525-2533. DOI: 10.12677/mos.2023.123232

Abstract

PM_{2.5} is one of the most harmful pollutants in air pollution, and accurate prediction of PM_{2.5} can provide a reliable basis for decision making in people's social activities. This paper proposes a PM_{2.5} prediction model, FE-LSTM (Feature Extraction-LSTM), with 24% fewer parameters to be trained than Auto-encoder, which can achieve high prediction accuracy but requires a large number of training parameters and increased computational resources. The FE-LSTM model is based on the SE attention mechanism, convolutional neural network and long short-term memory network modules. The initial features of the input tensor are first extracted using CNN, then the feature tensor is weighted by channel through SE attention, and the predicted pollution concentration is derived using LSTM after full concatenation and reconstruction. The FE-LSTM model and the comparison models were trained and tested on the Beijing PM_{2.5} dataset, and the results showed that the prediction accuracy of the FE-LSTM model proposed in this paper was better than that of the other comparison models.

Keywords

SE Attention, Convolutional Neural Network, Long Short-Term Memory Network, PM_{2.5} Prediction

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

长期处于高污染空气中的人容易患有哮喘、肺病甚至癌症等疾病，其中 PM_{2.5} 直径小，被人体吸入后可深入到肺泡，直接影响心肺功能。近年，我国已高度重视大气污染相关问题，针对不同地区制定了不同的减排目标，空气污染预测对于排查污染源、治理空气质量有着重大意义。可靠的大气污染预测模型可以提前给大众提供大气污染预警信息，警醒有关环保部门采取反制措施。

因此，很多专家和学者对大气污染预测开展了研究。例如，孙宁[1]等人使用 PSO-SVR 方法预测北京的 PM_{2.5} 浓度，利用 PSO 算法优化 SVR 模型中的相关参数，达到了比单个 SVR 模型更好的预测精度。Salina [2]等人结合了物联网技术和双向循环神经网络(RNN)模型实现了美国 PM_{2.5} 浓度的监测和预测系统。Dairi 等[3]开发了一个集成多方向注意力的变分自编码器预测模型(IMDA-VAE)，并在美国四个州的空气数据上进行了实验验证。张波[4]等人提出了一个 CAE-Learning 模型，先使用 CNN 模型学习到污染物和天气因素之间的关联特征，然后利用基于 LSTM 的自编码网络实现多站点的特征分析。宋飞扬[5]等人提出了一个基于时空特征的 KNN-LSTM 模型，考虑了多个监测站之间的区域关联，用于预测哈尔滨市的 PM_{2.5} 浓度，达到了比 BP 和 LSTM 模型更好的预测效果。Zhang [6]等人提出了一种基于自编码器和双向 LSTM 的深度学习模型来预测 PM_{2.5} 浓度，实验结果证明该自编码器的预测精度优于空气污染预测数值模型和几种机器学习模型。Qiao [7]等人提出了一个基于小波变换、自动编码器和长短期记忆网络(LSTM)的组合模型来预测多个城市的 PM₁₀ 浓度。Jin 等[8]提出了一个基于经验模态分解(EMD)，卷积神经网络(CNN)和门控循环单元结构(GRU)的组合模型来预测北京市的 PM_{2.5} 浓度。上述学者使用多种机器学习或深度学习方法来预测大气污染浓度，并取得了很好的预测精度，但是没有考虑到模型中需要训练的参数量，可能导致模型过大而无法部署在资源受限的设备上，应用前景不够广泛。

本文提出了一种基于特征提取和 LSTM 的 PM_{2.5} 浓度预测模型 FE-LSTM，在提升预测精度的同时，

尽量减少模型中需要训练的参数量, 试图花费较少的计算资源来获得更高的预测精度。该模型利用特征提取模块提取出输入张量的特征值, 然后输入到 LSTM 中进行预测。其中, 特征提取模块使用的是 CNN 和 SE-Attention 相结合的结构, 通过 CNN 将时间序列转化为三维特征张量, 然后使用 SE-Attention 加大重要特征的权重, 这有助于 LSTM 结构保留重要信息。

2. FE-LSTM 模型

本文提出的 FE-LSTM 模型结构如图 1 所示, 在 LSTM 之前加入特征提取模块, 在特征提取模块中使用 CNN 初步提取出输入序列的特征张量, 由于多核卷积生成的特征张量的不同通道会学习到不同的特征, 某些通道的特征对于后续预测不太重要, 为了使 LSTM 模型能够更好地学习到特征值, 使用 SE-Attention 分别计算初始特征张量中各个通道的权重, 并将归一化后的权重和原始特征张量相乘, 得到加权后的特征张量。再经过全连接和重新排列, 得到的新特征矩阵输入到 LSTM 结构中进行预测, 这有助于更重要的特征存入 LSTM 单元的长期细胞状态中, 从而提升模型精度。

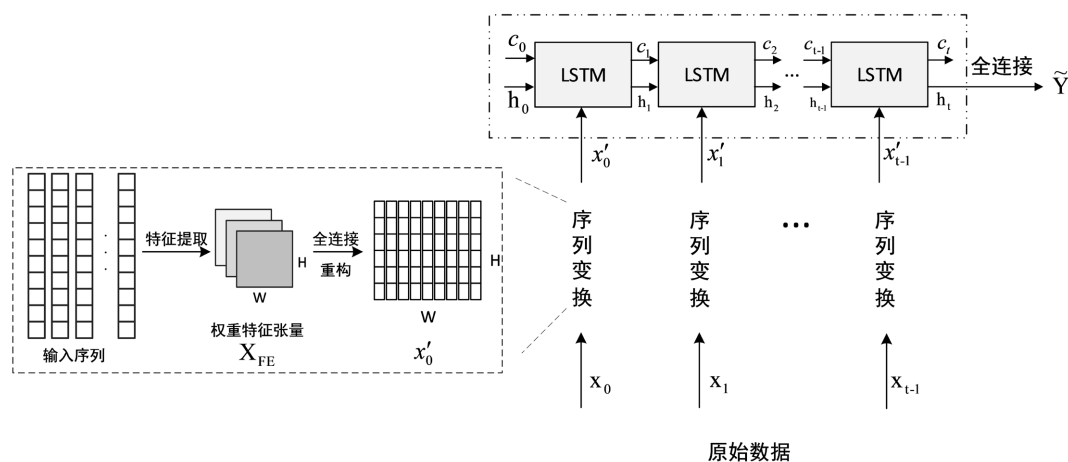


Figure 1. Structure of the FE-LSTM model

图 1. FE-LSTM 模型结构图

2.1. 特征提取

FE-LSTM 模型中使用特征提取模块进行序列变换, 其具体步骤如图 2 所示, 先使用卷积神经网络对输入序列进行多核卷积和池化后得到多个通道的初始特征张量, 然后使用 SE-Attention 在初始特征张量的通道尺度上进行加权处理, 得到权重特征张量, 再通过全连接将通道数减少为 1, 并重构为二维矩阵。

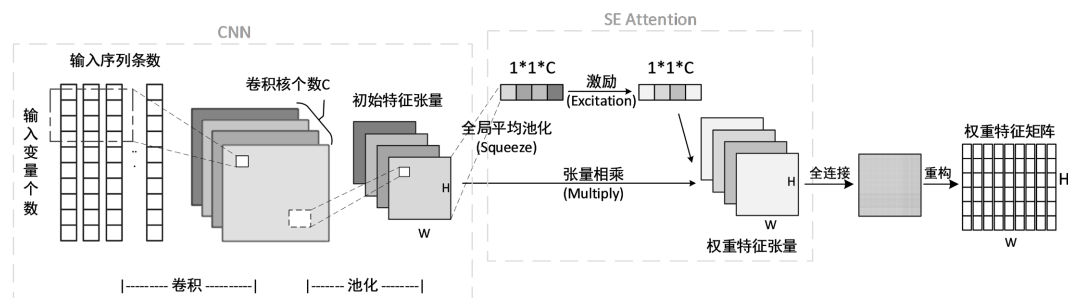


Figure 2. Structure of the feature extraction module

图 2. 特征提取模块结构示意图

1) 选用 CNN 提取输入序列的初始特征张量。卷积神经网络(CNN)是多层感知机的变种[9], 具有局部感知和权重共享的特性, 这有效减少了计算量, 减轻了训练时对设备算力的严苛要求。如图 2 所示, 本文使用的是 2D-CNN, 将输入序列看作通道数为 1 的二维矩阵, 卷积核的数量 C 即为卷积的次数, 每次卷积都能得到一个特征矩阵, 最终得到通道数为 C 的特征张量。卷积部分的计算公式如公式(1)所示。

$$X' = \delta_{relu}((K_C \otimes X) + b_C) \quad (1)$$

其中, X 为输入矩阵, X' 为提取出的特征张量, K_C 表示卷积核的权重系数, b_C 表示偏置向量。 \otimes 表示卷积操作, $\delta_{relu}(\dots)$ 表示激活层, 具体如公式(2)所示。

$$\delta_{relu}(z) = \begin{cases} 0, z < 0 \\ z, z \geq 0 \end{cases} \quad (2)$$

卷积神经网络中池化操作可以在不影响特征的本质的前提下减小特征矩阵的维度, 减少后续的计算量。在 FE-LSTM 模型中使用了 2×2 最大池化, 将卷积结果中每个 2×2 子矩阵中的最大值作为新矩阵的值, 对 C 个特征矩阵都进行上述池化降维操作后, 得到维度为 $W \times H \times C$ 的初始特征张量 X_p , 其中 W 、 H 分别表示初始特征张量的宽和高。

2) 选用 SE-Attention 对初始特征张量 X_p 的各个通道加权。SE 模块是由 Hu [10]等人提出, 通过学习通道间的依赖关系来增强特征。为了计算各个通道的权重, 先对每个通道的特征矩阵做全局平均池化(Squeeze)计算, 得到每个通道的平均值即维度为 $1 \times 1 \times C$ 的张量。其计算过程如公式(3)所示。

$$X_{sq} = \frac{1}{W \times H} \sum_{i=1}^H \sum_{j=1}^W X_p(i, j) \quad (3)$$

在激励(Excitation)的过程中, 使用了两次全连接层得到通道的权值, 第一次全连接压缩通道值至 $C/4$, 第二次全连接恢复通道数至原来的值 C , 先压缩再恢复能够减少需要训练的参数数量至原来的 $1/4$ 。使用 sigmoid 激励函数将权值映射到 $0 \sim 1$ 之间, 得到归一化后的通道权值 \tilde{X}_c 。其计算过程如公式(4)所示。

$$X_{sq} = \frac{1}{W \times H} \sum_{i=1}^H \sum_{j=1}^W X_p(i, j) \quad (4)$$

其中, W_s 、 b_s 和 W_r 、 b_r 分别表示两次全连接层的权重系数和偏置向量, σ_{sig} 计算过程如公式(5)所示。

$$\sigma_{sig}(z) = \frac{1}{1 + e^{-z}} \quad (5)$$

然后将待加权的初始特征张量 X_p 和权值 \tilde{X}_c 相乘, 将通道维度上的权重加入到特征张量中, 得到权重特征张量 X_{FE} 。

3) X_{FE} 为三维张量, 不能直接输入到 LSTM 网络中, 需要将其重构成二维矩阵。在此之前, 使用全连接层将通道数为 C 的特征张量 X_{FE} 映射成通道数为 1 的张量, 这能缩减重构后的二维矩阵的大小, 大大减少后续 LSTM 结构中需要训练的参数数量。最后将此维度为 $W \times H \times 1$ 的张量重构成维度为 $W \times H$ 的二维矩阵, 输入到 LSTM 结构中。

2.2. LSTM 结构

LSTM 最早由 Hochreiter 和 Schmidhuber [11]提出, 其单元结构[12]如图 3 所示。相比 RNN, LSTM 新增了遗忘门 σ_f 、输入门 σ_i 、输出门 σ_o 三个结构, 能够解决 RNN 容易出现梯度消失或梯度爆炸的问题, 将过去的信息存储在内部的细胞状态中, 随着新的输入递归更新。

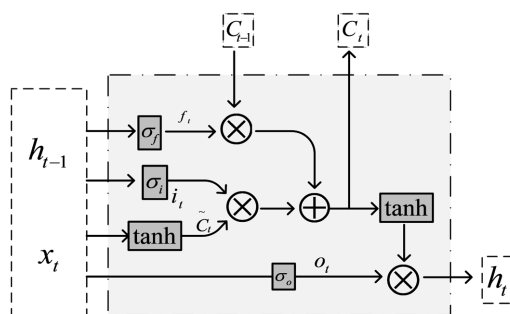


Figure 3. Structure of the LSTM unit
图 3. LSTM 单元结构示意图

图中， x_t 、 C_t 和 h_t 分别表示当前时刻的输入、细胞状态和输出， C_{t-1} 和 h_{t-1} 分别表示上一时刻的细胞状态和输出， σ 表示神经网络层，选用 Sigmoid 作为激励函数，其输出范围为 0~1，代表门限的开关。遗忘门、输入门和输出门的计算分别如公式(6)~(8)所示。

$$f_t = \sigma_f(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (6)$$

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

$$o_t = \sigma_o(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (8)$$

其中， W_f 、 W_i 和 W_o 分别表示遗忘门、输入门和输出门需要训练的权重矩阵， b_f 、 b_i 和 b_o 分别表示遗忘门、输入门和输出门需要训练的偏置向量。临时细胞状态 \tilde{C}_t 、当前细胞状态 C_t 和输出 h_t 的计算分别如公式(9)~(11)所示。

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (9)$$

$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t \quad (10)$$

$$h_t = o_t + \tanh(C_t) \quad (11)$$

2.3. 超参数选择

超参数需要在训练之前设置，本文使用网格搜索来确定最优的超参数组合，在 FE-LSTM 模型中，需要人为设置的超参数有：训练时的学习率 lr、批处理大小 batch_size、CNN 卷积核个数 filter、卷积核大小 kernel、卷积核的滑动步长 strider 以及 LSTM 的隐藏单元个数 hidden_size。经过实验比较，超参数取值如表 1 所示。

Table 1. Hyperparameter values chosen for the FE-LSTM model
表 1. FE-LSTM 模型选用的超参数值

lr	batch_size	filter	kernel	strider	hidden_size
0.0001	128	64	4	1	128

3. 实验结果与对比分析

3.1. 数据集介绍

为了验证提出模型的有效性，本文使用北京 PM_{2.5}数据集进行了测试。该数据集由 Liang [13]等人收集并在 UCI [14]数据集平台发布，包含 2010 年 1 月 1 日至 2014 年 12 月 31 日美国驻华大使馆附近的 PM_{2.5}

数据和北京首都国际机场的气象数据, 包括露点、温度、气压、风速、降雪量和降雨量等。其数据间隔为一个小时, 共 43,800 条数据。在实验中将此数据集划分为数据集、验证集和测试集, 数据比例分别为 0.7、0.15 和 0.15。

3.2. 模型评价指标

采用均方误差(MSE)、平均误差(MAE)、 R^2 系数用于评估模型性能。其中 MSE 和 MAE 越大说明模型效果越差, R^2 系数取值范围为 0~1 之间, 越接近 1 模型效果越好。它们的计算分别如公式(12)~(14)所示。

$$MSE = \frac{1}{N} \sum_{i=1}^N (\tilde{Y}^i - Y^i)^2 \quad (12)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |\tilde{Y}^i - Y^i| \quad (13)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (\tilde{Y}^i - Y^i)^2}{\sum_{i=1}^N (\bar{Y} - Y^i)^2} \quad (14)$$

其中, \tilde{Y}^i 是 i 时刻 $PM_{2.5}$ 的预测值, Y^i 是 i 时刻 $PM_{2.5}$ 的真实值, N 为测试集样本数量, \bar{Y} 表示 $PM_{2.5}$ 真实浓度的平均值。

3.3. 对比模型介绍

为了评估本模型在空气污染预测中的性能, 选取 LSTM, bi-LSTM, Auto-encoder 和 Cnn-AE 等模型进行了对比实验。

训练这四种基准模型时, 由于每个模型的收敛速度不一样, 在实验中使用早停法(Early Stopping)而非固定值。早停法能让模型得到充分训练的同时避免过拟合, 当 25 个 epoch 后模型在验证集上的误差仍没有下降的时候停止训练, 训练上限为 300 个 epoch。采用的是 Adam 优化器来优化深度学习模型, 损失函数被设置为均方误差(MSE)。

其余超参数使用网格搜索来选取, 其中, LSTM 神经元数量 hidden_size 从 16, 64, 128 中选取, 学习率 lr 从 0.01、0.001 和 0.0001 中选取, batch_size 从 64, 128, 256 中选取, 各个模型最终选用的超参数如下:

- 1) LSTM 模型选用简单的 LSTM 单元和全连接层的结构, LSTM 神经元数量取 128, batch size 为 128, 学习率选取 0.001。
- 2) Bi-LSTM 模型选用 Bi-LSTM 和全连接层的结构, LSTM 神经元数量取 128, batch size 为 128, 学习率选取 0.001。
- 3) Cnn-LSTM 模型, 输入数据通过 filters 为 64, 卷积核大小为 4, 步长为 1 的 2D 卷积层, 提取出的特征值输入到上述 LSTM 模型。
- 4) Auto-encoder 模型中编码器和解码器中都使用相同的 LSTM 结构, 分别从一层或两层, 单向或双向的 LSTM 单元中选取, 经过测试单层双向神经元数量 128 为最优结构, 并选取 batch size 为 128, 学习率选取 0.001。
- 5) Cnn-AE 模型, 输入数据通过 filters 为 64, 卷积核大小为 4, 步长为 1 的 2D 卷积层, 提取出的特征值输入到上述 Auto-encoder 模型。

3.4. 实验结果

在北京 $PM_{2.5}$ 数据集上, 使用本文提出的 FE-LSTM 模型和上述 5 个对比模型进行预测对比测试, 即

输入前十个小时的气象信息和 $PM_{2.5}$ 浓度来预测之后一个小时的 $PM_{2.5}$ 浓度, 实验结果如表 2 所示。从表 2 可以看出, 本文提出的 FE-LSTM 模型在北京 $PM_{2.5}$ 数据集上的 MSE 和 MAE 误差值均低于其他对比模型, R^2 相关系数均高于其他对比模型, FE-LSTM 模型相较于 LSTM、Bi-LSTM、Cnn-LSTM、Auto-encoder 和 Cnn-AE 在测试集上的 MAE 误差分别降低了 24.71%、22.06%、5.37%、20.41% 和 2.52%, MSE 误差分别降低了 21.60%、18.32%、19.52%、17.76% 和 14.38%, 这证明了 FE-LSTM 模型的预测值更贴近真实值, 预测精度更高。加入特征提取后的预测精度明显优于单一的 LSTM 模型, 也证明了本文设计的特征提取模块的有效性。

Table 2. Comparison of prediction results from multiple models on the Beijing $PM_{2.5}$ dataset

表 2. 在北京 $PM_{2.5}$ 数据集上各模型预测结果比较

模型名称	MSE	MAE	R2
LSTM	572.4357	14.8971	0.9077
Bi-LSTM	549.3633	14.3893	0.9115
Cnn-LSTM	557.6112	11.8519	0.9101
Auto-encoder	545.6932	14.0922	0.9121
Cnn-AE	524.1656	11.5053	0.9155
FE-LSTM	448.7654	11.2152	0.9277

各个模型中需要训练的参数量如表 3 所示, 包括权重和偏置向量等, 通常状况下, 参数量越大的深度学习模型需要更多的计算资源和时间来更新和优化, 会影响到模型的训练时间。在资源受限的设备上部署模型时也需要考虑到模型的参数量, 参数量较小的模型更容易在嵌入式设备上运行。

Table 3. The number of parameters to be trained for each model

表 3. 各模型需训练的参数量

模型名称	参数量	模型名称	参数量
LSTM	69,761	Auto-encoder	103,553
Bi-LSTM	139,521	Cnn-AE	856,385
Cnn-LSTM	296,641	FE-LSTM	79,234

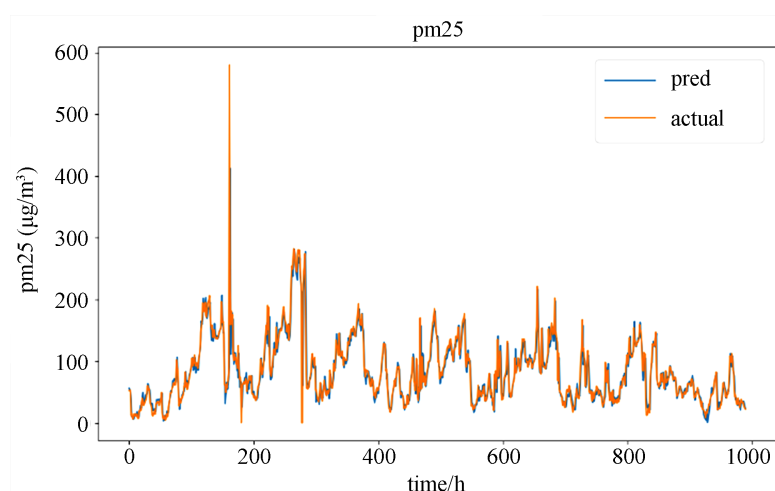


Figure 4. Visualisation of predicted and true values of the FE-LSTM model

图 4. FE-LSTM 模型预测值和真实值可视化

从表 3 可以看出 FE-LSTM 模型需要训练的参数只比 LSTM 增加 13.58%，比 Auto-encoder 减少了 23.48%，这说明 FE-LSTM 模型能够花费较少的计算资源就可以实现更为准确的预测结果。

FE-LSTM 模型在测试集前 1000 条数据上的 PM_{2.5} 浓度预测值和真实值的可视化曲线如图 4 所示，其中蓝色曲线为预测数据，橙色曲线为真实数据。从图 4 中可以看出，本文提出的模型预测效果较好，能够预测出 PM_{2.5} 浓度的变化趋势，但是在峰值和谷值等数据突变处表现一般。

4. 结论与展望

本文提出了一个基于特征提取和 LSTM 的 PM_{2.5} 浓度预测模型，先使用卷积神经网络提取出输入序列的初始特征张量，再使用 SE-Attention 机制得到带有权重信息的加权特征张量，然后输入到 LSTM 结构中得到预测结果。在北京 PM_{2.5} 数据集的测试结果表明，本文提出的模型能很好地预测出 PM_{2.5} 浓度值的变化趋势，与其余五个基准模型相比，FE-LSTM 预测出的数据和真实值更为贴近，相关系数均高于其余几个模型，均方误差和平均误差均小于其余模型，预测精度更高。FE-LSTM 模型能够在确保精度的前提下降低需训练的参数量，表明该模型耗费的计算资源更少，便于部署在资源受限的设备上，能够在轻量级的嵌入式设备上运行，具有一定的应用价值。

本文提出的 FE-LSTM 模型仍有提升空间，由图 4 的预测曲线可以看出本模型在突变处的预测效果不太理想，之后可以继续尝试将更前沿的先进算法加入模型中，例如引入对突变数据敏感度更高的相关算法，来获得更高的预测精度。

基金项目

浙江省公益技术研究项目(2015C31024)。

参考文献

- [1] 孙宁, 陈田, 徐桂安. 粒子群优化支持向量回归算法的空气污染情况预测[J]. 上海电机学院学报, 2022, 25(1): 7-11+16.
- [2] Salinas, D., Flunkert, V., Gasthaus, J. and Januschowski, T. (2020) DeepAR: Probabilistic Forecasting with Autoregressive Recurrent Networks. *International Journal of Forecasting*, **36**, 1181-1191. <https://doi.org/10.1016/j.ijforecast.2019.07.001>
- [3] Dairi, A., Harrou, F., Khadraoui, S., et al. (2021) Integrated Multiple Directed Attention-Based Deep Learning for Improved Air Pollution Forecasting. *IEEE Transactions on Instrumentation and Measurement*, **70**, 1-15. <https://doi.org/10.1109/TIM.2021.3091511>
- [4] 张波, 陆云杰, 秦东明, 等. 一种卷积自编码深度学习的空气污染多站点联合预测模型[J]. 电子学报, 2022, 50(6): 1410-1427.
- [5] 宋飞扬, 铁治欣, 黄泽华, 丁成富. 基于 KNN-LSTM 的 PM_{2.5} 浓度预测模型[J]. 计算机系统应用, 2020, 29(7): 193-198. <https://doi.org/10.15888/j.cnki.csa.007490>
- [6] Zhang, B., Zhang, H. and Zhao, G. (2020) Constructing a PM_{2.5} Concentration Prediction Model by Combining Auto-Encoder with Bi-LSTM Neural Networks. *Environmental Modelling & Software*, **124**, Article ID: 104600. <https://doi.org/10.1016/j.envsoft.2019.104600>
- [7] Qiao, W., Wang, Y. and Zhang, J. (2021) An Innovative Coupled Model in View of Wavelet Transform for Predicting Short-Term PM10 Concentration. *Journal of Environmental Management*, **289**, Article ID: 112438. <https://doi.org/10.1016/j.jenvman.2021.112438>
- [8] Jin, X.B., Yang, N.X. and Wang, X.Y. (2020) Deep Hybrid Model Based on EMD with Classification by Frequency Characteristics for Long-Term Air Quality Prediction. *Mathematics*, **8**, 214. <https://doi.org/10.3390/math8020214>
- [9] Qin, D., Yu, J., Zou, G., et al. (2019) A Novel Combined Prediction Scheme Based on CNN and LSTM for Urban PM_{2.5} Concentration. *IEEE Access*, **7**, 20050-20059. <https://doi.org/10.1109/ACCESS.2019.2897028>
- [10] Hu, J., Shen, L. and Sun, G. (2018) Squeeze-and-Excitation Networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 7132-7141.

-
- <https://doi.org/10.1109/CVPR.2018.00745>
- [11] 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>
- [12] Graves, A. (2012) Long Short-Term Memory. In: *Supervised Sequence Labelling with Recurrent Neural Networks*, Vol. 385, Springer, Berlin, Heidelberg, 37-45. https://doi.org/10.1007/978-3-642-24797-2_4
- [13] Liang, X., Zou, T., Guo, B., *et al.* (2015) Assessing Beijing's PM_{2.5} Pollution: Severity, Weather Impact, APEC and Winter Heating. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, **471**.
<https://doi.org/10.1098/rspa.2015.0257>
- [14] UCI 北京数据集[Z]. <http://archive.ics.uci.edu/ml/datasets/Beijing+PM2.5+Data>