

基于时空特征提取的空气污染物 PM_{2.5} 预测

凌德森, 王晓凯

(山西大学 物理电子工程学院, 太原 030006)

摘要: 为了充分挖掘多因素数据间的时空特征信息, 解决在多种因素相互影响下不能准确预测 PM_{2.5} 值的问题, 提出了一种融合了局部加权回归的周期趋势分解 (STL, seasonal-trend decomposition procedure based on loess) 算法、卷积长短期记忆网络 (ConvLSTM, convolutional long short-term memory network) 和门控循环单元 (GRU, gated recurrent unit) 的 PM_{2.5} 预测方法; 首先利用 STL 算法将 PM_{2.5} 数据进行分解, 将分解得到的序列分别与其他因素相融合; 搭建 ConvLSTM-GRU 模型, 并利用贝叶斯寻优算法进行超参数寻优; 将融合数据传入 ConvLSTM 网络中进行时空特征提取, 再将提取后的特征序列传入 GRU 网络中进行预测; 通过与 ConvLSTM-GRU 模型、CNN-GRU 模型以及 GRU 模型的预测结果进行比较实验, 证明所提模型具有误差小、预测效果好等特点。

关键词: 卷积长短期记忆网络; 门控循环单元; 贝叶斯寻优算法; 时空特征

Prediction of Air Pollutant PM_{2.5} Based on Time-space Feature Extraction

LING Desen, WANG Xiaokai

(College of Physics and Electronic Engineering, Shanxi University, Taiyuan 030006, China)

Abstract: In order to fully mine the spatiotemporal feature information between multi-factor data, and solve the problem that PM_{2.5} value cannot be accurately predicted under the influence of multiple factors, a PM_{2.5} prediction method is proposed to combine with a seasonal-trend decomposition procedure based on Loess (STL) algorithm, convolutional long short-term memory network (ConvLSTM) and gated recurrent unit (GRU). Firstly, the STL algorithm is used to decompose the PM_{2.5} data and fuse the decomposed sequence with other factors; The ConvLSTM-GRU model is built, and the Bayesian optimization algorithm is used to search for super parameters; The fused data is transferred to extract the ConvLSTM network for time-space feature, and then the extracted feature sequence is transferred to the GRU network for the prediction. Compared with the prediction results of the ConvLSTM-GRU model, CNN-GRU model and GRU model, the results show that the proposed model has the characteristics of small error and good prediction effect.

Keywords: ConvLSTM; GRU; bayesian optimization algorithm; time-space feature

0 引言

经过蓝天保卫战的努力, 我国城市的细颗粒物 (PM_{2.5}) 已经下降了许多, 但在北方的很多城市到了秋冬季后, 雾霾现象依然很严重, 严重影响了人们的日常生活, 而雾霾的主要成分为细颗粒物 (PM_{2.5}), 因此对 PM_{2.5} 值的预测对空气污染防治、守护蓝天保卫成果具有重要意义。

目前, 国内外的学者在 PM_{2.5} 预测方面已有了大量的研究^[1], 主要的方法有数据驱动方法^[2], 如 Rui Zhao 等^[3]

利用多元线性回归 (MLR, multiple linear regression) 模型来实现 PM_{2.5} 的短期预测, Liu Wei 等^[4] 利用粒子群算法优化支持向量机 (SVM, support vector machine) 对 PM_{2.5} 等级进行预测, 之后随着人工神经的发展, 深度学习作为机器学习的一个分支, 是经典机器学习的一个进步。与经典机器学习不同, 经典机器学习仍然需要人工辅助来提取特征, 深度学习具有自动学习原始数据特征的优点。如贾佳美等^[5] 利用 BP 神经网络对 PM_{2.5} 日均浓度进行预测; Jeya^[6] 等利用双向 LSTM 模型预测 PM_{2.5} 污染物浓

收稿日期: 2023-01-04; 修回日期: 2023-03-07。

基金项目: 山西省重点研发计划(高新技术领域)(201803D121102)。

作者简介: 凌德森(1995-), 男, 硕士研究生。

通讯作者: 王晓凯(1963-), 男, 博士, 教授。

引用格式: 凌德森, 王晓凯. 基于时空特征提取的空气污染物 PM_{2.5} 预测[J]. 计算机测量与控制, 2023, 31(11): 31-37.

度, 并采用均方根误差、平均绝对误差以及平均绝对百分比误差进行误差评估分析, 最终结果表明此模型相较于对比模型精度得到了较大的提升。另外, 混合预测模型也是最近几年重点研究对象, 如 Bai 等^[7]利用改进的极限学习机结合离群点检测和校正算法、数据分解策略和多目标优化算法, 形成了 PM2.5 浓度的混合模型, 蒋洪讯等^[8]提出了一种基于小波分解 (WD, wavelet decomposition) 的长短期记忆网络 (LSTM, long short-term memory network) 用来预测 PM2.5 的浓度, 结果表明数据经过分解处理后, 能够降低数据噪声的影响, 提高模型的预测性能。为了提高模型的预测精度, 学者还利用模型对数据的隐藏特征进行提取, 如 Jin 等^[9]利用经验模态分解 (EMD, empirical mode decomposition) 将 PM2.5 数据进行分解, 并将分解后的数据根据频率分量进行分组, 最后利用 CNN-GRU 模型进行预测, 实验结果表明利用卷积神经网络 (CNN, convolution neural network) 网络进行特征提取之后会进一步提高模型的预测精度, 卫晓旭等^[10]提出一种多元预测模型, 先利用 CNN 网络对多因素数据进行空间特征提取, 再使用双向长短期记忆 (BiLSTM, bidirectional long short-term memory network) 网络对多因素数据中包含的时间特征进行挖掘并预测, 进一步提高了模型对污染物浓度预测的精度。但是该模型先对数据的空间特征进行提取, 再对时间特征进行挖掘, 这可能会导致在空间特征提取后数据中某些时间特征被忽略的问题。

目前, 虽然很多的模型也都考虑了多因素之间的相互影响^[11-13], 但多数的研究中只是偏向于对数据的时间特征信息或空间特征信息的挖掘, 并没有同时对时间特征和空间特征进行提取, 如前文介绍的 CNN 网络, 虽然相较于单一模型有了很大的提升, 但模型对数据特征的提取不够完善, 不能很好的挖掘数据的隐藏信息, 也是阻碍预测精度提升的原因。因此本文提出了一种时空特征提取的 PM2.5 浓度预测模型方法, 为了能够更容易获取预测数据的内部特征, 先利用 STL 算法对预测序列进行分解, 再将分解后的序列混合着其他污染物数据和大气因素数据传入 ConvLSTM 网络进行时间特征和空间特征进行同时提取, 之后传入到利用贝叶斯优化后的 GRU 网络进行预测。结果表明, 本文提出的模型具有更高的精度。

1 研究方法的基本原理

1.1 ConvLSTM 网络

ConvLSTM 是双重组合的深度学习方法, 它是 LSTM 架构的混合变体, 使用卷积算子而不是矩阵乘法来进行状态输入和状态到状态转换^[14]。这使算法能够处理时空数据并使用本地邻居的输入和先前的状态来确定网格中特定单元格的即将到来的状态^[15]。其运行机制如下:

$$i_t = \sigma(W_{xi} * X_t + W_{hi} * H_{t-1} + W_{ci} \circ C_{t-1} + b_i) \quad (1)$$

$$f_t = \sigma(W_{xf} * X_t + W_{hf} * H_{t-1} + W_{cf} \circ C_{t-1} + b_f) \quad (2)$$

$$C_t = f_t \circ C_{t-1} + i_t \circ \tanh(W_{xc} * X_t + W_{hc} * H_{t-1} + b_c) \quad (3)$$

$$o_t = \sigma(W_{xo} * X_t + W_{ho} * H_{t-1} + W_{co} \circ C_t + b_o) \quad (4)$$

$$H_t = o_t \circ \tanh(C_t) \quad (5)$$

式 (1) ~ (5) 中: * 和 \circ 分别表示卷积和 Hadamard 乘积, i , f 和 o 分别表示输入门、遗忘门和输出门, H 表示时间序列的隐藏状态, C 表示每个细胞的输出, σ 为激活函数, W 表示连接状态之间的权重。

1.2 STL 分解算法

STL 分解算法是一种常见的时间序列分解算法, 它通过局部加权回归 (Loess)^[16]将时间序列分解为趋势项、季节项和随机波动的残差项, 如式 (6) 表示。

$$x_t = T_t + S_t + R_t, t = 1, 2, \dots, N \quad (6)$$

其中: x_t 、 T_t 、 S_t 和 R_t 分别表示在 t 时刻的时间序列观测值、趋势项、季节项和残差项。另外 STL 可分为内循环和外循环, 其中内循环主要用来进行趋势拟合和周期分量的计算, 内循环的主要步骤如下。

步骤 1: 赋初值 $T_t^{(0)} = 0$, 去趋势 $x_t - T_t^{(k)}$;

步骤 2: 子序列平滑, 将每个子序列进行 LOESS 过程, 前后各延展 1 个时间点, 组合得到长度为 $(N + 2n_p)$ 的时间序列 C_t^{k+1} ;

步骤 3: 低通滤波, 将步骤 2 得到的时间序列进行 3 次的滑动平均, 再进行 1 次 LOESS 过程, 得到长度为 N 的时间序列 L_t^{k+1} , 去除周期性差异;

步骤 4: 将步骤 2 得到的时间序列进行趋势项去除, 为:

$$S_t^{k+1} = C_t^{k+1} - L_t^{k+1} \quad (7)$$

步骤 5: 季节项去除 $x_t - S_t^{k+1}$;

步骤 6: 进行趋势平滑, 将步骤 5 得到的序列进行 LOESS 得到 T_t^{k+1} ;

步骤 7: 判断步骤 6 得到的时间序列是否收敛, 若不收敛重新执行步骤 1 到步骤 6, 直到时间序列收敛, 若收敛则内循环结束。

外循环通过引入稳健权重来处理内循环中出现的异常残差值, 可以减少迭代过程中异常值的影响, 从而提高了算法的鲁棒性^[17]。

1.3 贝叶斯优化

贝叶斯优化是由 Pelikan 等^[18]学者在 1998 年提出的, 它在已知有限样本条件下, 通过构造黑箱函数输出的后验概率来寻找函数的最优值, 其主要目标是在超参数空间内寻找到泛化性能最优的多维超参数 x^* , 公式如下:

$$x^* = \underset{x \in X}{\operatorname{argmax}} f(x) \quad (8)$$

式中, $f(x)$ 是超参数向量 x 到模型泛化性能的映射。由于用黑箱函数评估模型代价高昂且复杂, 为了克服此缺点, 常常利用高斯过程模型作为代理模型来评估模型, 改进后的模型可以考虑之前的参数信息, 不断地更新先验知识^[19]。在超参数优化的问题中, 贝叶斯优化是当前为数不多的, 具有较好的收敛性理论保证的超参数估计方法^[20]。

1.4 GRU 网络

GRU 网络是 Chung 等^[21] 在 2014 年提出的一种新的 RNN 变体循环神经网络 (RNN) 的一种变体, 是在 LSTM 网络上的进一步的改进, 它能够在降低了细胞结构复杂度的同时, 保证和 LSTM 的模型精度相当的同时, 提高网络的运行效率, 使得网络在训练中能够更快的收敛^[22]。其前向传公式如下:

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t]) \quad (9)$$

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) \quad (10)$$

$$\tilde{h}_t = \tanh(W_{\tilde{h}} \cdot [r_t * h_{t-1}, x_t]) \quad (11)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \quad (12)$$

式 (9) ~ (12) 中, x_t 为当前输入, h_{t-1} 为前一时间刻传递出的隐藏状态信息, r_t 表示当前时刻重置门的状态值, z_t 表示当前时刻更新们的状态值, σ 是非线性激活函数, \tilde{h}_t 表示要进行选择的隐藏状态集, $*$ 表示矩阵乘积, h_t 为当前时刻的隐藏状态信息, $W_r, W_z, W_{\tilde{h}}$ 为所要学习的 3 个权重向量。

2 实验方法与实验数据

2.1 实验方法

为了使模型具备时空特征提取的能力, 本文采用 ConvLSTM 模型进行时空特征提取, 将提取后的特征再利用 GRU 网络进行预测。此外, 在加入多种特征因素后, 直接使用空气污染物原始数据进行预测, 会降低模型的预测精度, 因此采用 STL 分解算法先将待预测的空气污染物序列进行分解, 再进行预测。模型结构如图 1, 具体步骤如下。

步骤 1: 为了解决各种数据输入量纲不一致的问题, 这里将要预测的污染物浓度序列、相关的污染物浓度序列, 以及气象因素数据进行归一化处理, 原始数据主要是空气污染物浓度数据。又因原始数据的分布改变可能会造成空气污染物浓度分布信息的丢失, 为了保证归一化处理后的数据不改变原始数据的分布情况, 因此本文利用 Max-Min 方法对数据作归一化处理。

$$y_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (13)$$

式中, y_i 为归一化后的数据, x_i 为原始数据, x_{\max}, x_{\min} 分别为原始数据中的最大值、最小值。

步骤 2: 将归一化后的待预测的 PM2.5 污染物浓度序列, 采用 STL 算法进行分解得到季节性序列、趋势序列和残差序列, 能够有效降低序列中噪声的影响, 并更有利于提高后续对数据特征的提取。

步骤 3: 将步骤 2 得到的 3 种子序列分别加入归一化后的其他污染物浓度序列及气象因素, 并分别利用 ConvLSTM 模型进行时空特征提取, 时空特征的提取能够更加突出对预测效果影响更大的因素的影响, 提高模型预测的准确率。

步骤 4: 利用步骤 3 提取到的特征值作为 GRU 网络

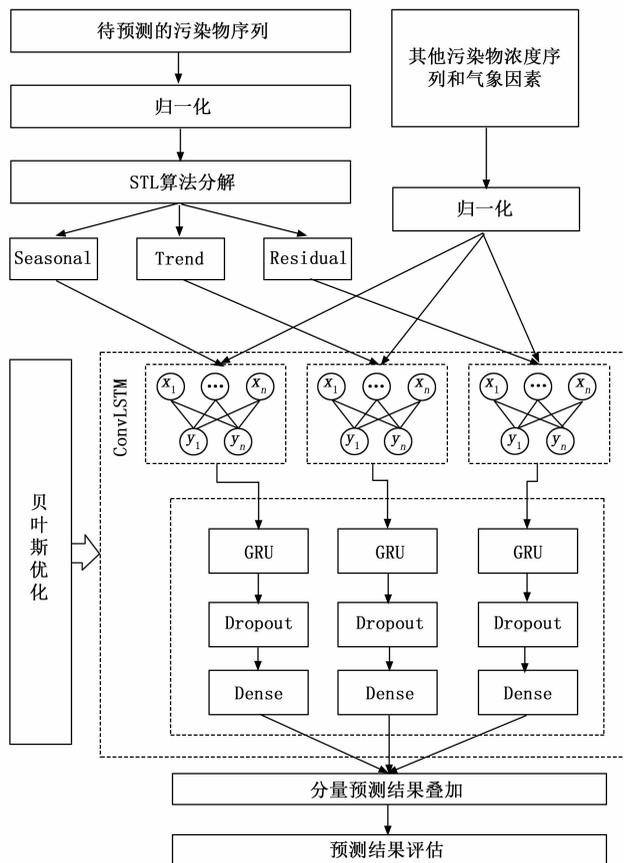


图 1 STL-ConvLSTM-GRU 模型的整体结构

的输入值, 预测结构的误差值作为目标函数, 使用贝叶斯算法对 GRU 网络进行超参数优化。

步骤 5: 利用步骤 4 优化好的 GRU 预测模型对步骤 3 处理后的子序列分别进行预测, 为了降低模型过拟合的问题, 在 GRU 层后增加 Dropout 层, Dropout 正则化方法能够有效降低模型过拟合的问题。步骤 6: 将子序列预测结果通过 Dense 层进行输出, 将最后的子序列预测结果叠加, 并进行反归一化操作得到最终的预测结果。

2.2 数据来源与相关性分析

本文选择以某市 2017 年 1 月 1 日至 2017 年 7 月 31 日每小时采样一次的历史实时空气污染物浓度数据为研究样例, 数据集收集于中国环境监测总站的全国城市空气质量实时发布平台 (<https://air.cnemc.cn>) 和真气网 (<https://www.aqistudy.cn/>) 发布的数据^[23], 其中包括 PM10、PM2.5、SO₂、NO₂、O₃ 和 CO 6 项主要污染物浓度以及温度、湿度、风级、风向 4 种气象数据。

由于空气污染物浓度和气象数据收集过程十分不易, 因此造成收集到的数据部分缺失, 对于缺失的数据分别按照污染物和气象数据来进行补缺, 对于空气污染物的缺失值采用相邻数据取均值的方法, 即将空缺处的污染物数据利用其上一时刻和下一时刻的值进行平均运算得出, 气象数据则直接使用 0 值来补充。补充后的各污染物浓度数据

如图 2 所示。

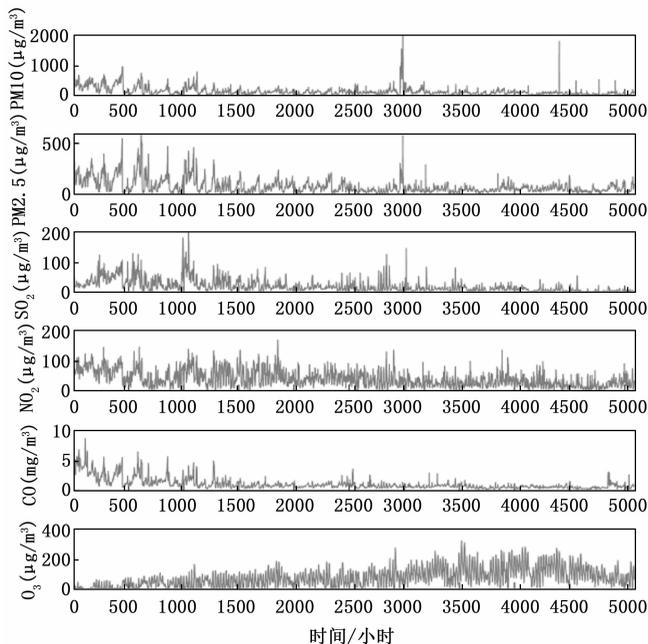


图 2 六种空气污染物浓度的真实数据

从图 2 中可以直观的看出六种空气污染物浓度之间的波动具有一定的相关性，其中 PM2.5、PM10、SO2、NO2、CO 五种空气污染物浓度的变化表现出一种同增同减的正相关性，而 O3 与其它 5 种空气污染物浓度之间表现出一种一减一增的负相关性。为了更加深入的研究六种空气污染物之间存在的相关性，本文利用皮尔逊 (Pearson) 相关系数进行相关性研究。

Pearson 相关系数是英国数学家卡尔·皮尔逊在弗朗西斯·高尔顿的研究基础之上演变而来。Pearson 相关系数可以用来表示两个变量之间的相关性，反应两个变量之间变化趋势的方向以及程度^[24]。若有两个长度为 n 的序列 $\{x_1, x_2, \dots, x_i, \dots, x_n\}$ 和 $\{y_1, y_2, \dots, y_i, \dots, y_n\}$ ，则 Pearson 相关系数 r 的计算公式为：

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (14)$$

式 (14) 中， \bar{x} 和 \bar{y} 分别为两个序列集合的平均值， r 的取值范围为 $[-1, 1]$ 。当 r 的取值在 $[-1, 0)$ 时，两个序列之间呈现负相关性；当 r 的取值为 0 时，两个序列之间不存在相关性；当 r 的取值在 $(0, 1]$ 时，两个序列之间存在正相关性。根据 Pearson 相关系数公式，得到如图 3 所示的空气污染物之间的相关热力图。

从空气污染物浓度的 Pearson 相关系数图 (图 3) 中可以看出，PM2.5、PM10、SO2、NO2、CO、O3 六种空气污染物浓度之间相互存在着一定的相关性，其中 PM10、PM2.5、SO2、NO2、CO 之间相互呈现不同程度的正相关

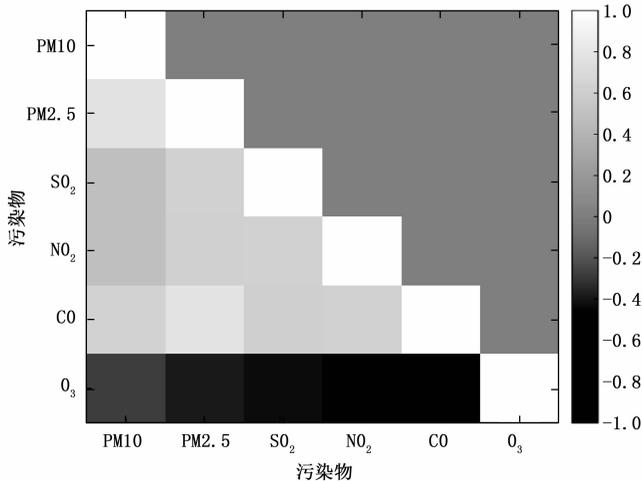


图 3 六种空气污染物浓度之间的相关系数

性，O3 和其他 5 种空气污染物之间呈现不同程度的负相关性。

2.3 实验设备

为了保证所有实验模型的公平性，本次所有实验均在 Windows11 系统下，利用 Python3.6、Matlab R2019b 以及 tensorflow2.0 平台搭建和运行实验模型。

3 实验过程与结果分析

3.1 STL 分解

为了能够更加准确地获取时间序列的内部特征，将待预测的 PM2.5 浓度时间序列进行 STL 分解，经过多次实验后，将 STL 分解周期设置为 230 小时，得到 PM2.5 的季节分量、反映 PM2.5 变化趋势的趋势项、以及反映 PM2.5 随机波动的残余项。从图 4 中可以看出虽然 PM2.5 数据的波动比较混乱，但是仍然具有一定的周期性规律，从分解后的趋势项来看从 1 月到 7 月的这段时间内随着气温的逐渐升高，PM2.5 浓度逐渐下降并慢慢趋于平稳，这种现象表明

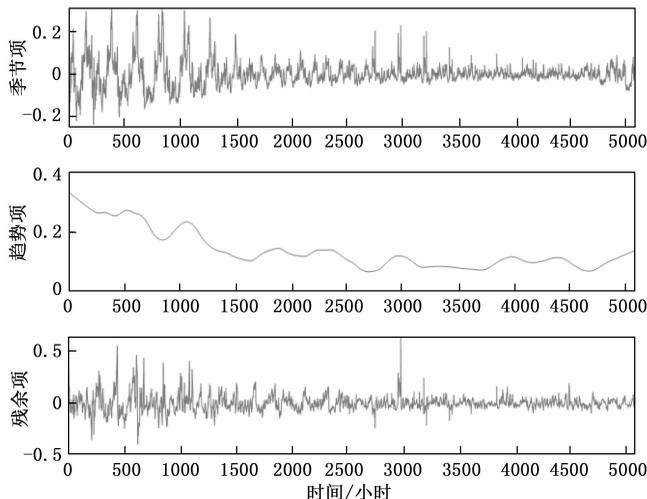


图 4 经 STL 分解后的 PM2.5 浓度

雾霾等 PM2.5 污染现象的出现可能和冬季取暖消耗的资源较多有一定的联系。

3.2 模型参数的选择

在模型构建过程中需要设置模型的参数, 参数的选择也是模型构建的重点, 合适的参数关系到模型的预测误差和泛化能力^[25]。在实际应用的过程中 ConvLSTM 网络卷积核的熟练关系到模型对特征提取的完整程度, 而 GRU 网络中神经元的数量、学习率的大小、正则化率的大小、迭代次数都影响着模型的最终预测结果, 因此设置合适的超参数是十分重要的。

以往在超参数选择中无论是根据自己的经验来不断地尝试, 还是利用网格搜索法来进行多种尝试, 都不仅耗时耗力, 还会因为个人经验的积累程度影响到最终的结果。根据 1.3 节的分析可知, 贝叶斯优化算法能够较大范围的搜索超参数的范围, 减少模型陷入局部最优的情况。因此本文采用贝叶斯寻优算法来降低人力的消耗, 减少尝试法以及网格搜索法带来的个人经验的影响。贝叶斯寻优步骤如下:

1) 设定这 5 个参数的选择范围, 以及每次搜索参数的步长;

2) 设置初始化寻优参数, 利用控制变量的方法, 先固定一个超参数的大小, 依次对其它超参数进行遍历, 建立不同的预测模型;

3) 每次建立的模型都会计算一次模型运行的误差值, 对比每一次误差值的大小, 选择出这一组最优的超参数。

4) 调整之前固定参数的大小, 再次对其它超参数进行遍历, 根据 3) 的描述再次选择出一组最优的超参数;

5) 当所有的选择参数都进行遍历寻优之后, 对比每一组最优超参数下的误差值, 选择出计算误差最小的一组作为模型的最优超参数。

因为经过 STL 分解后的 PM2.5 数据有三项, 所有将同一模型根传入数据的不同, 分别进行超参数优化。贝叶斯寻优的最终结果如表 1 所示。

表 1 贝叶斯寻优结果

超参数	贝叶斯寻优结果		
	季节项	趋势项	残余项
卷积核数	64	64	64
神经元数	128	64	140
Dropout	0.2	0.1	0.2
学习率	0.001	0.001	0.001
迭代次数	70	40	60

3.3 评价指标

为了能够更好的衡量模型的预测效果, 本文采用了 3 个评价指标: 平均绝对误差 (MAE, mean absolute error)、均方根误差 (RMSE, root mean squared error) 和平均绝对百分比误差 (MAPE, mean absolute percentage error)

来分析预测结果与真实值之间偏差。一般来说 RMSE、MAE 和 MAPE 值越小则预测结果与真实值之间的偏差越小, 模型预测效果越好;

$$RMSE(y_i, \hat{y}_i) = \sqrt{\frac{1}{n} \left[\sum_{i=1}^n (y_i - \hat{y}_i)^2 \right]} \quad (15)$$

$$MAE(y_i, \hat{y}_i) = \frac{1}{n} \left[\sum_{i=1}^n |y_i - \hat{y}_i| \right] \quad (16)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \quad (17)$$

式 (15) ~ (17) 中, y_i 是真实的空气污染物浓度值, \hat{y}_i 为空气污染物浓度的预测值。

3.4 实验结果分析

将讨论的 PM2.5 数据经 STL 算法分解成季节项、趋势项和残余项 3 个子序列, 如 3.1 节分析, 并将 3 个子序列分别与其他 5 种污染物数据以及温度、湿度、风级、风向 4 种气象数据的前 4 882 条数据作为训练集、后 206 条数据作为测试集, 预测步长选择 8 小时, 即使用前的 8 个小时的数据预测预测后 1 个小时的值。模型预测结果如图 5 所示。

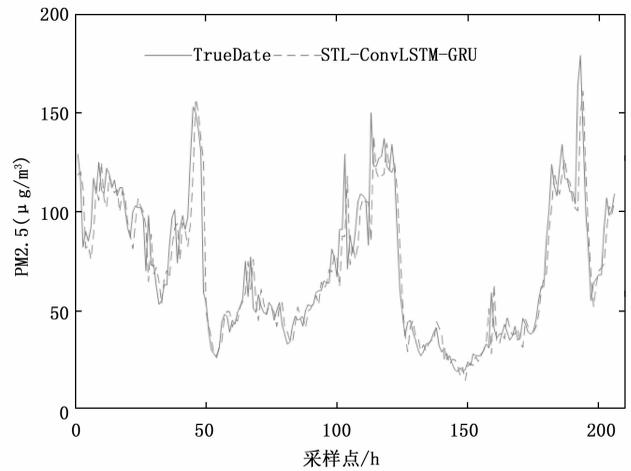


图 5 模型预测结果

从图 5 可以看出, STL-ConvLSTM-GRU 模型的 PM2.5 浓度预测结果与实际 PM2.5 浓度的整体趋势较为贴合, 符合 PM2.5 浓度的变化波动趋势。虽然在某些 PM2.5 浓度变化波动复杂的时间段内, 该模型的预测结果和实际 PM2.5 浓度有一些误差, 但从 PM2.5 浓度的变化趋势上看, 预测结果与实际值也是相符合的, 表明该模型能够较为准确的预测 PM2.5 浓度值, 并且在 PM2.5 浓度变化复杂的情况下也能很好的预测 PM2.5 的波动趋势。

3.5 对比实验

为了验证本文提出模型的优劣性, 除了本文主要的研究模型外, 还搭建了 ConvLSTM-GRU 模型、常见的 CNN-GRU 模型以及 GRU 模型作为实验对比模型。

从 GRU 网络的预测结果图 (图 6) 可以看出, 在相同模型参数的情况下, GRU 网络对 PM2.5 浓度的预测结果

与实际值的偏离度较大，甚至在一些 PM2.5 浓度波动较为复杂的情形下，基本失去了预测 PM2.5 浓度趋势的可能性。

将 GRU 网络的预测结果与 ConvLSTM-GRU 模型和 CNN-GRU 模型的 PM2.5 浓度预测结果 (图 7、8) 进行对比可以发现，虽然 ConvLSTM-GRU 模型和 CNN-GRU 模型的预测结果离真实值都有着一定的差距，但是通过比较这 3 种模型预测结果与实际值的拟合程度，可以得出在这 3 种模型中仅仅使用单一 GRU 模型进行预测的效果最差，与实际走势的偏离度最高，预测值与实际值的偏差最大，而另外两种模型虽然也与真实值有所差距，但预测结果与实际值的偏离度却有着明显的改善，因此可以得出在对 PM2.5 浓度预测前对多因素数据进行特征提取，能够有效的提高模型对 PM2.5 浓度的预测准确率。

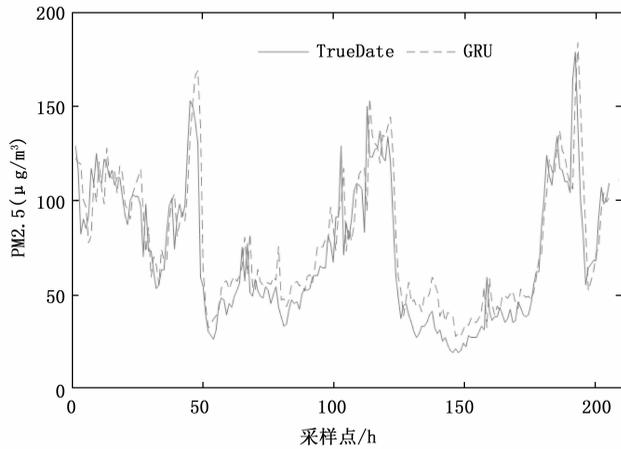


图 6 GRU 网络的预测结果

通过对比 ConvLSTM-GRU 模型和 CNN-GRU 模型的 PM2.5 浓度预测结果，发现采用 ConvLSTM 进行时空特征提取之后，对于 PM2.5 浓度波动比较频繁的部分以及浓度趋势变化较尖锐的时刻，预测结果更加贴近真实 PM2.5 浓度的走势，再通过观察表 2 中各评价指标值，可以得出 ConvLSTM-GRU 模型 RMSE、MAE 以及 MAPE 的值要小于使用 CNN 进行特征提取的 CNN-GRU 模型，因此可以得出 ConvLSTM-GRU 模型的预测效果要优于 CNN-GRU 模型，也即是使用 ConvLSTM 进行时间和空间的特征提取后，比仅仅利用 CNN 网络进行空间特征提取更加有效的提高了模型的预测精度。

通过对比 STL-ConvLSTM-GRU 模型和 ConvLSTM-GRU 模型的预测结果和结果评估表可以看出，STL-ConvLSTM-GRU 模型的预测结果要更加优于 ConvLSTM-GRU 模型的预测结果，这表明使用 STL 分解算法对 PM2.5 数据进行分解，能够更有效的突出数据中隐藏的特征信息、降低噪声对 PM2.5 浓度预测模型的影响，并进一步提高模型的预测准确度。

通过对比 4 种模型的 RMSE、MAE 和 MAPE 三种评价

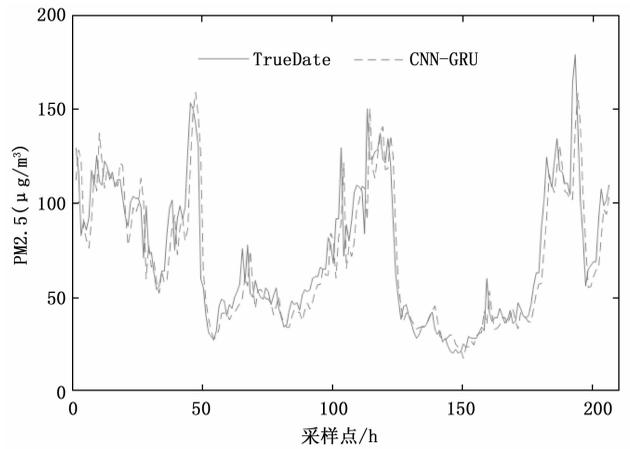


图 7 CNN-GRU 模型的预测结果

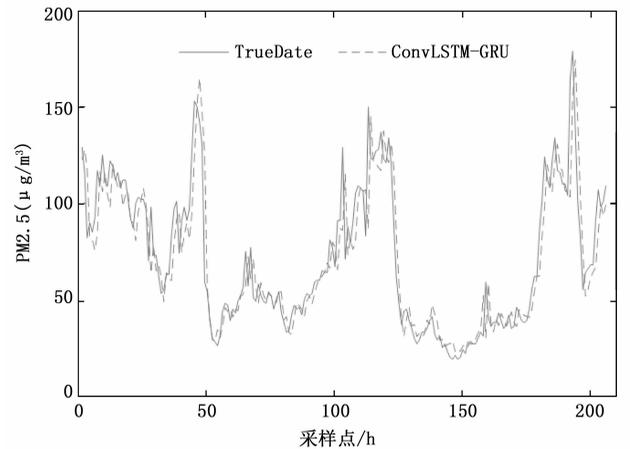


图 8 ConvLSTM-GRU 模型的预测结果

指标，STL-ConvLSTM-GRU 模型相较于其它 3 种模型，RMSE 分别降低了 5.5%、12.54%、16.54%；MAE 分别降低 8.2%、18.05%、28.82%，MAPE 也有着不同程度的降低。从以上的分析可以得出，STL-ConvLSTM-GRU 模型的预测结果相较于其它 3 种模型效果更优，整体结果更加贴近真实的 PM2.5 浓度的变化趋势。

表 2 4 种预测模型性能对比

预测模型	评价指标		
	RMSE	MAE	MAPE(%)
STL-ConvLSTM-GRU	14.22	9.21	13.14
ConvLSTM-GRU	15.05	9.97	14.07
CNN-GRU	16.26	11.20	15.53
GRU	17.04	12.94	22.45

4 结束语

1) 相较于单一的 GRU 模型，经过 ConvLSTM 进行时空特征提取后的 GRU 模型预测效果更好，因此采用 ConvLSTM 对 PM2.5 数据进行时空特征提取具有可行性。ConvLSTM-GRU 模型比 GRU 模型的 RMSE、MAE、MAPE

分别降低了 11.67%、22.95%、8.38%。

2) 通过对比 ConvLSTM-GRU 模型和 CNN-GRU 模型的预测结果, 可以得出采用 ConvLSTM 对数据进行时间和空间的特征提取能够更多的挖掘出 PM_{2.5} 浓度数据中的隐藏信息, 提高模型的预测技能高度。

3) 由于多变量预测的复杂性较高, 为了更好的提取内部特征信息, 利用 STL 算法先将待预测的 PM_{2.5} 数据分解, 再和其它影响因素一起利用 ConvLSTM 进行时空特征提取, 相较于 ConvLSTM-GRU 模型, STL-ConvLSTM-GRU 模型更进一步提高了预测精度, 证明了本文所研究的模型优越性。

4) 本文只是将影响因素一并传入到预测模型中, 并没有考虑影响因素与待预测序列之间相关度大小的影响, 因此后面将对数据之间相关度大小的影响进行实验。

参考文献:

[1] 王磊, 杨翠丽, 乔俊飞. 基于回声状态网络的 PM_{2.5} 预测研究 [J]. 控制工程, 2019, 26 (1): 1-5.

[2] 刘辉, 邓达华. 一种基于深度学习的当日 PM_{2.5} 混合集成预测方法 (英文) [J]. Journal of Central South University, 2022, 29 (6): 2074-2083.

[3] ZHAO R, GU X, XUE B, et al. Short period PM_{2.5} prediction based on multivariate linear regression model [J]. PloS one, 2018, 13 (7): e0201011.

[4] LIU W, GUO G, CHEN F, et al. Meteorological pattern analysis assisted daily PM_{2.5} grades prediction using SVM optimized by PSO algorithm [J]. Atmospheric Pollution Research, 2019, 10 (5): 1482-1491.

[5] 贾佳美, 池凯凯, 吴哲翔. 改进粒子群优化 BP 神经网络的 PM_{2.5} 预测 [J]. 计算机工程与设计, 2021, 42 (12): 3495-3501.

[6] JEYA S, SANKARI L. Air pollution prediction by deep learning model [C] //2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS). IEEE, 2020: 736-741.

[7] BAI L, LIU Z, WANG J. Novel hybrid extreme learning machine and multi-objective optimization algorithm for air pollution prediction [J]. Applied Mathematical Modelling, 2022, 106: 177-198.

[8] 蒋洪迅, 闫超超, 张立峰. 基于时序分解和神经网络的 PM_{2.5} 浓度预测研究——以沈阳市为例 [J]. 系统科学与数学, 2021, 41 (12): 3446-3460.

[9] JIN X B, YANG N X, WANG X Y, et al. Deep hybrid model based on EMD with classification by frequency characteristics for long-term air quality prediction [J]. Mathematics, 2020, 8 (2): 214.

[10] 卫晓旭, 王晓凯, 朱涛, 等. 基于多元分解的大气污染深

度学习预测方法 [J]. 计算机仿真, 2021, 38 (5): 467-471, 483.

- [11] ELANGASINGHE M A, SINGHAL N, DIRKS K N, et al. Complex time series analysis of PM₁₀ and PM_{2.5} for a coastal site using artificial neural network modeling and K-means clustering [J]. Atmospheric Environment, 2014, 94: 106-116.
- [12] XU Z, XIA X, LIU X, et al. Combining DMSP/OLS nighttime light with echo state network for prediction of daily PM_{2.5} average concentrations in Shanghai, China [J]. Atmosphere, 2015, 6 (10): 1507-1520.
- [13] ZOU B, WANG M, WAN N, et al. Spatial modeling of PM_{2.5} concentrations with a multifactorial radial basis function neural network [J]. Environmental Science and Pollution Research, 2015, 22 (14): 10395-10404.
- [14] MOISHIN M, DEO R C, PRASAD R, et al. Designing deep-based learning flood forecast model with ConvLSTM hybrid algorithm [J]. IEEE Access, 2021, 9: 50982-50993.
- [15] SHI X, CHEN Z, WANG H, et al. Convolutional LSTM network: A machine learning approach for precipitation nowcasting [J]. Advances in neural information processing systems, 2015, 802-810.
- [16] 王晓燕, 郎贺, 王品, 等. 基于 STL 分解的平均故障间隔时间组合预测 [J]. 机床与液压, 2021, 49 (17): 196-200.
- [17] 赵建东, 朱丹, 刘佳欣. 基于时间序列分解与门控循环单元的地铁换乘客流预测 [J]. 华南理工大学学报 (自然科学版), 2022, 50 (5): 22-31.
- [18] PELIKAN M, GOLDBERG D E, CANTÚ - PAZ E. BOA: The Bayesian optimization algorithm [C] //Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-99), 1999: 525-532.
- [19] 李亚茹, 张宇来, 王佳晨. 面向超参数估计的贝叶斯优化方法综述 [J]. 计算机科学, 2022, 49 (S1): 86-92.
- [20] 江敏. 贝叶斯优化算法的若干问题研究及应用 [D]. 上海: 上海大学, 2012.
- [21] CHUNG J, GULCEHRE C, CHO K H, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling [J]. arXiv preprint arXiv: 1412.3555, 2014.
- [22] 刘超, 雷晨, 李树刚, 等. 基于 CNN-GRU 的瓦斯浓度预测模型及应用 [J]. 中国安全生产科学技术, 2022, 18 (9): 62-68.
- [23] 凌德森, 王晓凯, 朱涛. 基于自适应时序分解的空气污染物浓度预测 [J]. 测控技术, 2023, 42 (1): 83-91, 98.
- [24] 程娟娟. 高校科研与教学关系实证研究——基于皮尔逊相关系数的分析 [J]. 中国高校科技, 2022, 410 (10): 46-52.
- [25] 孟升卫, 冯晓晓, 庞景月, 等. 基于 LS-SVM 和核密度估计的概率性风电功率预测 [J]. 计算机测量与控制, 2019, 27 (12): 34-38.