

融合深度神经网络特征的 ARIMAX 雾霾 PM2.5 预测

张钰洁¹, 王 钰², 杨杏丽¹

(1. 山西大学 数学科学学院, 山西 太原 030006;

2. 山西大学 现代教育技术学院, 山西 太原 030006)

摘 要:目前,雾霾污染问题是关乎国计民生的重大问题,它已经对人们的生产、生活、身体健康,以及生态环境和气候变化都产生了很大的影响。这样,如何通过监测雾霾变化获取的雾霾相关信息去准确预测雾霾污染物的浓度,以防治和减轻雾霾造成的严重后果变得尤为重要。因此,通过在简单有效的传统 ARIMAX 模型基础上融入深度神经网络语义特征,提出了一种新的雾霾 PM2.5 浓度预测框架。首先,把对雾霾预测有显著影响的气象因子温度、压力、相对湿度数据转换为图像数据;然后,运用 ResNet-50 (Residual Network-50) 卷积神经网络模型提取深度语义特征,进而运用主成分分析 (Principal Component Analysis, PCA) 技术处理高维特征,得到最佳深度神经网络特征组合;最后,用 ARIMAX 技术建立雾霾 PM2.5 浓度预测模型。在收集的山西省 2015~2019 年 PM2.5 浓度和气象因子数据集上验证了该预测框架在皮尔逊相关系数 (Pearson's Correlation Coefficient, PCC)、均方误差 (Mean Square Error, MSE)、均方根误差 (Root Mean Squared Error, RMSE) 和平均绝对误差 (Mean Absolute Error, MAE) 度量下,对于 1、3、5 和 7 天长期预测,都始终优于传统的简单差分自回归滑动平均 (Autoregressive Integrated Moving Average, ARIMA) 模型、三因素 ARIMAX 模型、多元回归模型、ResNet-多元回归模型、长短期记忆网络 (Long and Short-Term Memory, LSTM) 模型和支持向量机 (Support Vector Machine, SVM) 模型。

关键词:PM2.5 预测; ARIMAX 模型; ResNet 神经网络; 主成分分析技术; 深度语义特征

中图分类号: TP183; TP181

文献标识码: A

文章编号: 1673-629X(2023)02-0167-06

doi: 10.3969/j.issn.1673-629X.2023.02.025

Haze PM2.5 Concentration Prediction Based on ARIMAX Model with Deep Convolutional Neural Network Features

ZHANG Yu-jie¹, WANG Yu², YANG Xing-li¹

(1. School of Mathematical Sciences, Shanxi University, Taiyuan 030006, China;

2. School of Modern Educational Technology, Shanxi University, Taiyuan 030006, China)

Abstract: At present, the problem of haze pollution is a major issue related to the national economy and people's livelihood, which has already had a great impact on people's production, life, health, and ecological environment and climate change. In this way, how to accurately predict the concentration of haze pollution by monitoring the haze-related information obtained by monitoring the haze changes in order to prevent and reduce the serious consequences of haze has become particularly important. Therefore, by incorporating the semantic features of deep neural networks on the basis of the ARIMAX model, we propose a new haze PM2.5 concentration prediction framework. Firstly, the numerical meteorological data (temperature, pressure and relative humidity that have a significant impact on haze prediction) is transformed into image data. Secondly, the ResNet-50 (Residual Network-50) model is used to extract deep semantic features and PCA (Principal Component Analysis) is to process high-dimensional features for the best combination of deep neural network features. Finally, ARIMAX technology is used to establish a haze PM2.5 concentration prediction model. Furthermore, on the collected data set of PM2.5 concentration and meteorological factors in Shanxi Province from 2015 to 2019, under the Pearson's Correlation Coefficient (PCC), Mean Square Error (MSE), Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) measurements, the experimental results demonstrate the proposed prediction frameworks are always superior to the traditional

收稿日期: 2022-04-03

修回日期: 2022-08-05

基金项目: 山西省应用基础研究计划项目 (201901D111034); 国家自然科学基金项目 (62076156); 统计与数据科学前沿理论及应用教育部重点实验室开放基金 (KLATASDS2007)

作者简介: 张钰洁 (1997-), 女, 硕士研究生, 研究方向为深度学习、统计机器学习; 通信作者: 王 钰 (1981-), 男, 教授, 博士, CCF 会员 (G6976M), 研究方向为统计机器学习、图像处理。

Autoregressive Integrated Moving Average (ARIMA), three-factor ARIMAX, multiple regression, ResNet-multiple regression, Long and Short-Term Memory (LSTM) and Support Vector Machine (SVM), for 1, 3, 5 and 7-day long and short-term prediction.

Key words: PM2.5 prediction; ARIMAX; ResNet; principal component analysis; deep semantic feature

0 引言

雾霾天气是一种大气污染状态,雾霾是对大气中各种悬浮颗粒物含量超标的笼统表述。雾霾的主要成分包括 PM2.5 颗粒物、PM10 颗粒物、臭氧、二氧化氮、二氧化硫和一氧化碳等,尤其是由有毒物质如重金属和致癌物质等组成的有机化合物 PM2.5 (空气动力学当量直径小于等于 2.5 微米的颗粒物)被认为是造成雾霾天气的“元凶”^[1-2]。

PM2.5 对人类健康的危害极大,会导致心血管和心肺部疾病、呼吸系统问题,增加癌症的发病几率,也可能造成婴儿的畸形发育等^[3-4]。另外,PM2.5 雾霾污染对气候变化和生态环境也有很大影响,PM2.5 浓度的增加可能是极端天气和气候事件增加的主要原因之一,如引起城市大气酸雨、光化学烟雾现象,导致大气能见度下降,极大阻碍空中、水面和陆面交通的正常运行。

因此,准确监测雾霾变化,获取雾霾相关信息,尤其是及时准确预报雾霾污染物的浓度,对于大气科学、大气环境监测、气候分析、气象预报、环境保护、国民经济和军事等诸多方面均具有重要意义^[5-6]。

目前,文献中已经提出了多种雾霾污染浓度的预测模型和技术,如差分自回归滑动平均(Autoregressive Integrated Moving Average, ARIMA)模型^[7]、多元线性回归模型^[8]、季节性差分自回归滑动平均(Seasonal Autoregressive Integrated Moving Average, SARIMA)模型^[9]、向量自回归模型^[10]、支持向量机(Support Vector Machine, SVM)模型^[11]、决策树^[12]、随机森林^[13-14]、长短期记忆网络(Long and Short-Term Memory, LSTM)^[15-17]等。进一步,朱旭辉等^[18]针对单个 SVM 方法的不鲁棒问题,通过使用多数投票算法对多个 SVM 模型进行选择集成,给出了一种集成的雾霾天气预测方法,实验结果表明集成后的 SVM 方法具有更高的精度和鲁棒性。刘梦炀等^[19]构建了一种基于 LSTM 和全连接神经网络的混合神经网络模型,并提出了数据桶划分的训练方式来解决由于训练数据与预测数据存在较长时间间隔导致精度下降的问题,进而实现 PM2.5 浓度的精确预测。然而,上述方法都仅仅使用了雾霾数据本身的信息,但实际上温度、压力、相对湿度等气象因子对于雾霾的预测都有显著的影响。为此,该文考虑融合温度、压力、相对湿度的气象数据和雾霾数据给出更为准确的雾霾预测结果。

然而发现,简单融合三个常用气象因子的

ARIMAX 雾霾预测模型并不能得到满意的预测结果。这是因为雾霾的形成和发展过程是非常复杂的,这些模型可能无法充分有效地模拟复杂的雾霾过程,从而无法捕获重要的深度语义特征,造成预测结果不准确。

特别地,深度神经网络模型由于其强大的特征表示能力得到了极大的关注,为此,该文考虑将深度卷积神经网络模型提取的深度语义特征融入到 ARIMAX 模型,给出一种新的雾霾序列预测框架。首先,由于卷积神经网络^[20]能够自动准确地学习图像空间相关性等深度语义信息,将数值气象因子数据转化为图像数据;然后,基于具有优越性能的 ResNet-50 (Residual Network-50)卷积神经网络模型^[21]提取气象因子图像序列中的深层特征,进一步采用主成分分析(Principal Component Analysis, PCA)技术处理高维特征,得到最佳深度语义特征组合;最后,用简单有效的 ARIMAX 模型建模捕获雾霾的时间序列相关信息。

1 提出的雾霾预测框架

1.1 ARIMAX 模型

ARIMAX 模型是指带回归项的 ARIMA 模型,又称扩展的 ARIMA 模型,通过回归项的引入进一步提高模型的预测效果。此模型把输出序列表示为随机波动的过去值和预测因子序列的过去值的结合,即:

$$\begin{cases} Y_t = \mu + \sum_{i=1}^k \frac{\Theta_i(B)}{\Phi_i(B)} B^{l_i} X_{it} + \varepsilon_t \\ \varepsilon_t = \frac{\Theta(B)}{\Phi(B)} a_t \end{cases}$$

其中, Y_t 表示输出序列, $\{X_{1t}\}, \{X_{2t}\}, \dots, \{X_{kt}\}$ 表示输入变量序列, $\{\varepsilon_t\}$ 为回归残差序列, $\{a_t\}$ 为零均值白噪声序列, B 为滞后操作算子, $\Theta_i(B)$ 为第 i 个输入变量的自回归系数多项式, $\Phi_i(B)$ 为第 i 个输入变量的滑动平均系数多项式, l_i 为第 i 个输入变量的滞后阶数,且:

$$\Phi(B) = 1 - \varphi_1 B - \dots - \varphi_p B^p$$

$$\Theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$$

1.2 ResNet-50 神经网络模型

ResNet-50 模型是目前深度学习中的主流卷积神经网络模型,其内部的残差块使用了跳跃连接,极大地缓解了梯度消失问题,克服了由于网络深度加深而产生的学习效率变低与准确率无法有效提升的问题。因此,该文使用 ResNet-50 模型提取深度语义图像特征。具体地,ResNet-50 模型的网络结构如表 1 所列,它包含了 49 个卷积层、1 个全连接层,总共由 16 个残差块

组成,如图 1 所示,每个残差块包含 3 个卷积层、1 个跳转连接和激活函数。经过前五部分的卷积计算,输出 $7 \times 7 \times 2\,048$ 的特征图,然后池化层将其转化成一个 2 048 维特征向量。

表 1 ResNet-50 结构

layer name	output size	50-layer
conv1	$112 \times 112 \times 64$	$7 \times 7, 64, \text{stride } 2$ $3 \times 3 \text{ max pool, stride } 2$
conv2_x	$56 \times 56 \times 256$	$\begin{bmatrix} 1 \times 1 & 64 \\ 3 \times 3 & 64 \\ 1 \times 1 & 256 \end{bmatrix} \times 3$
conv3_x	$28 \times 28 \times 512$	$\begin{bmatrix} 1 \times 1 & 128 \\ 3 \times 3 & 128 \\ 1 \times 1 & 512 \end{bmatrix} \times 4$
conv4_x	$14 \times 14 \times 1\,024$	$\begin{bmatrix} 1 \times 1 & 256 \\ 3 \times 3 & 256 \\ 1 \times 1 & 1\,024 \end{bmatrix} \times 6$
conv5_x	$7 \times 7 \times 2\,048$	$\begin{bmatrix} 1 \times 1 & 512 \\ 3 \times 3 & 512 \\ 1 \times 1 & 2\,048 \end{bmatrix} \times 3$
	$1 \times 1 \times 2\,048$	Average pool, fc, softmax

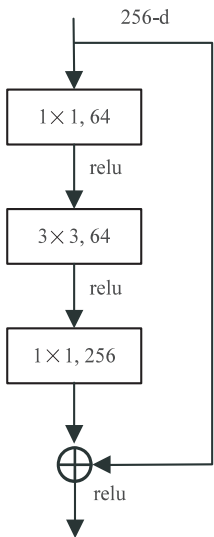


图 1 残差块网络结构

1.3 提出方法

提出的预测框架如图 2 所示。

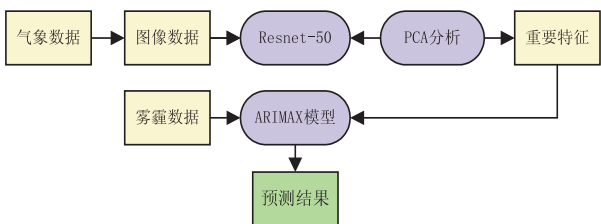


图 2 预测框架

首先,将三因素气象因子数值型数据转化为三通

道气象因子图像,即根据目标城市和邻近城市的经纬度等相关信息生成目标城市和邻近城市的地图,并用归一化后的三个基本气象因子数值数据填充相应城市地图的像素,生成三通道气象因子图,其中温度、气压和相对湿度各对应 t 时刻图像的一个通道。继而运用 ResNet-50 模型提取三通道气象因子图像深度语义特征,每个图像的特征均为 2 048 维。由于图像的特征维度较高,可能对后续的预测性能带来不好的影响。因此,用 PCA 技术进行降维,使用 MSE 评价准则选取主成分的最优个数,该个数记为 n^* ,使用前 n^* 个主成分得到最优的特征组合,即 $\{X_{1t}\}, \{X_{2t}\}, \dots, \{X_{n^*t}\}$ 。该组合包含了数据气象因子的大部分信息,作为后续 ARIMAX 模型的 X_t , PM2.5 序列数据作为 ARIMAX 模型中的 Y_t 。由于 ARIMAX 模型要求输入的时间序列数据是平稳序列,因此在建立模型前检验 $\{X_{1t}\}, \{X_{2t}\}, \dots, \{X_{n^*t}\}$, PM2.5 序列的平稳性和方差齐性。若数据是平稳的,则运用最优特征组合和 PM2.5 数据建立 ARIMAX 模型实现 PM2.5 的高精度多步预测。若数据不平稳,则对数据进行差分变换至平稳序列再建立模型。具体地,融合深度卷积神经网络特征的 ARIMAX 雾霾 PM2.5 浓度预测算法过程如下:

- 输入:三因素气象数据、PM2.5 数据;
- a) 数值数据转化为三通道图像
 - b) ResNet-50 提取图像特征,特征维数为 2 048
 - c) for $n = 1$ to 2 048
 - PCA 提取前 n 个主成分
 - 建立 ARIMAX 模型
 - 计算 MSE_n
 - End
 - d) 选取最小 MSE 对应的主成分个数 $n^* = \operatorname{argmin}_n (MSE_n)$
 - e) 将前 n^* 个主成分 $\{X_{1t}\}, \{X_{2t}\}, \dots, \{X_{n^*t}\}$ 作为最佳特征组合
 - f) 检验 $\{X_{1t}\}, \{X_{2t}\}, \dots, \{X_{n^*t}\}$, PM2.5 序列的平稳性
 - g) 如果序列是平稳的,则转到第 i 步,否则第 h 步
 - h) 对数据进行(一次或多次)差分操作
 - i) 检验方差齐性,如果序列是异方差,转到第 j 步,否则转到第 k 步
 - j) 对原始数据进行数据变换,转到第 g 步
 - k) 识别模型,确定 p 和 q
- 输出:预测值。

2 实验分析

为了验证提出的雾霾预测框架的有效性,收集了一个数据量大小为 1 826 的 PM2.5 浓度和气象因子数据集,并在此数据集上将提出的预测框架与广泛使用的 ARIMA 模型、三因素 ARIMAX 模型、多元回归模

型、ResNet-多元回归模型、LSTM 模型和 SVM 模型在四个常用的性能评价指标上进行了对比。

2.1 数据及其分析

主要收集了山西省 11 个城市(太原、大同、朔州、忻州、阳泉、吕梁、晋中、长治、晋城、临汾、运城)2015 年 1 月至 2019 年 12 月的空气质量和气象数据:(a)主要空气污染源 PM2.5 浓度数据;(b)三个基本气象因子:温度、气压和相对湿度数据。然后,将三个基本气象因子数值数据转化为图像数据。具体地,根据目标城市和邻近城市的经纬度等相关信息生成目标城市和邻近城市的地图,并用归一化后的三个基本气象因子数值数据填充相应城市地图的像素,生成 1 826 张的三通道气象因子图像,如图 3 所示。

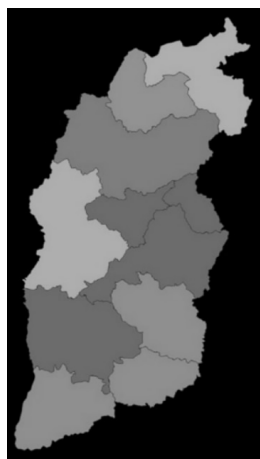


图 3 三通道气象因子图

基于生成的图像,通过 ResNet-50 模型提取 2 048 维特征,继而为了保证模型的预测精度,采用 PCA 降维,找到最佳的特征组合。经过实验,如图 4 所示,因子个数为 34 时得到了最小的 RMSE,即 34 维特征为 PCA 提取的最佳特征组合。最后,基于上述 34 维重要深度语义特征和 PM2.5 序列,经过平稳性、方差齐性和白噪声检验后,建立 ARIMAX 模型预测 PM2.5 浓度。例如,在大同市地区,大同市气象和雾霾序列数据通过平稳性、方差齐性和白噪声检验后,建立 ARIMAX(2,1,3)模型,对大同市未来 PM2.5 浓度进行预测。

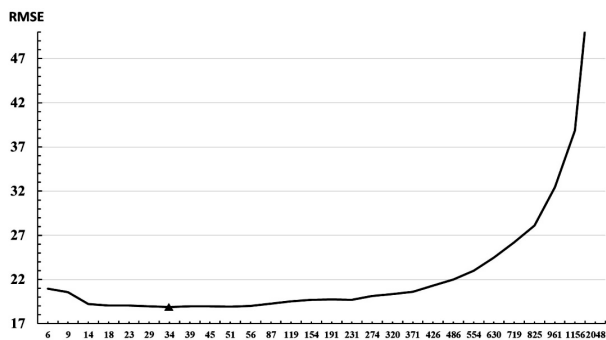


图 4 特征个数与 RMSE 变化关系

2.2 评价指标

使用几种常用的雾霾污染预测评价指标评估预测模型性能:皮尔逊相关系数(Pearson's Correlation Coefficient, PCC)、均方误差(Mean Square Error, MSE)、均方根误差(Root Mean Squared Error, RMSE)和平均绝对误差(Mean Absolute Error, MAE),它们的表达式如下:

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2}$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i|$$

$$PCC = \frac{\sum_{i=1}^n Y_i \hat{Y}_i}{\left(\sum_{i=1}^n Y_i^2 \right) \left(\sum_{i=1}^n \hat{Y}_i^2 \right)}$$

其中, Y_i 是真实值, \hat{Y}_i 是预测值。

2.3 实验结果

将该文提出的方法与 LSTM 模型、SVM(三因素)模型、多元回归(三因素)模型、ResNet-多元回归模型、ARIMA 模型、ARIMAX(三因素)模型进行比较:

(1) LSTM 模型:原始 LSTM 模型,直接使用 PM2.5 序列建立模型。

(2) SVM(三因素)模型:直接使用三个气象因子作为 PM2.5 预测的影响因素,建立 SVM 回归模型。

(3) 多元(三因素)模型:直接使用三个气象因子作为 PM2.5 预测的影响因素,建立多元回归模型。

(4) ResNet-多元回归模型:首先使用 ResNet 神经网络提取三通道气象因子图深度语义特征,经过主成分分析找到最佳特征组合,继而基于最佳特征组合建立多元回归模型。

(5) ARIMA 模型:仅使用 PM2.5 序列建立 ARIMA 模型。

(6) ARIMAX(三因素)模型:直接使用三个气象因子作为 PM2.5 预测的影响因素,建立 ARIMAX 模型。

表 2 ~ 表 5 给出了 11 个城市的平均 MSE、RMSE、MAE 和 PCC 值对于 1、3、5、7 天的长短期 PM2.5 预测的多个模型的对比结果,图 5 给出了预测长度为 1 天时,各个城市的多个模型的 MSE、RMSE、MAE 和 PCC 对比结果。

首先,在 1、3、5、7 天长短期 PM2.5 预测下,提出的预测框架在四个指标下始终优于 ARIMA 模型。例如,如表 2 所列,在预测时长为 1 天时,11 个城市的平均 MSE、RMSE、MAE 和 PCC 值,ARIMA 模型分别为

992.043、30.978、21.486、0.632,而提出的预测框架分别为 327.135、17.275、12.958、0.769,显著优于 ARIMA 模型。这些结果验证了单纯使用 ARIMA 模型预测 PM2.5 浓度是远远不够的,ARIMA 模型无法考虑到影响雾霾浓度的因素,导致预测结果不准确。

表 2 其他模型与提出框架对四种度量的
1 天 PM2.5 浓度预测比较

Algorithm	MSE	RMSE	MAE	PCC
LSTM	1 328.243	35.751	25.751	0.494
SVM(三因素)	667.948	25.084	18.779	0.660
多元(三因素)	801.386	27.121	22.374	0.592
ResNet-多元	487.670	21.202	16.411	0.736
ARIMA	992.043	30.978	21.486	0.632
ARIMAX(三因素)	375.118	18.691	13.478	0.716
Ours	327.135	17.275	12.958	0.769

表 3 其他模型与提出框架对四种度量的
3 天 PM2.5 浓度预测比较

Algorithm	MSE	RMSE	MAE	PCC
LSTM	1 593.332	38.453	28.111	0.266
SVM(三因素)	680.334	25.317	18.972	0.656
多元(三因素)	816.911	27.388	22.622	0.582
ResNet-多元	496.975	21.389	16.596	0.731
ARIMA	1 478.992	37.804	26.178	0.453
ARIMAX(三因素)	512.659	21.600	15.596	0.605
Ours	397.343	18.729	13.814	0.730

表 4 其他模型与提出框架对四种度量的
5 天 PM2.5 浓度预测比较

Algorithm	MSE	RMSE	MAE	PCC
LSTM	1 482.855	37.315	27.798	0.148
SVM(三因素)	690.291	25.507	19.128	0.652
多元(三因素)	823.795	27.513	22.724	0.575
ResNet-多元	502.980	21.510	16.695	0.726
ARIMA	1 747.397	40.670	25.874	0.447
ARIMAX(三因素)	552.812	22.318	16.163	0.571
Ours	420.608	19.156	14.091	0.717

接着,鉴于单纯使用 ARIMA 模型预测 PM2.5 浓度模型性能较差,该文将影响雾霾浓度的三个气象因子加入到 ARIMA 模型,建立了三因素 ARIMAX 模型。在 1、3、5、7 天长短期 PM2.5 预测下,提出的预测框架在四个指标下始终优于三因素 ARIMAX 模型。在某些情况下,提出的预测框架改进达到了 10%~30%。例如,如表 5 所列,在 7 天 PM2.5 浓度预测的情况下,所提出的预测框架分别与三因素 ARIMAX 模型的 MSE、RMSE、MAE 和 PCC 度量进行比较,性能提高了

27.9%、16.5%、15.5%、27.2%。这是因为,雾霾的形成和发展过程是非常复杂的,而直接将三因素数据建模可能无法充分有效地模拟复杂的雾霾过程,从而导致模型性能较低。这一原因在多元回归中也得到了验证。例如,在四个粒度下,使用卷积神经网络提取的特征作为自变量建立的多元回归模型的性能均远远超过直接用三个气象因素建立的多元回归模型的性能。

表 5 其他模型与提出框架对四种度量的
7 天 PM2.5 浓度预测比较

Algorithm	MSE	RMSE	MAE	PCC
LSTM	1 207.862	34.275	25.365	0.233
SVM(三因素)	697.241	25.640	19.242	0.651
多元(三因素)	834.076	27.688	22.883	0.569
ResNet-多元	490.301	21.289	16.536	0.719
ARIMA	1 990.426	43.272	26.607	0.410
ARIMAX(三因素)	571.094	22.628	16.412	0.556
Ours	411.504	18.886	13.868	0.707

然后,该文使用卷积神经网络提取的特征作为自变量建立多元回归模型,实验结果验证了单独利用卷积提取图像特征做预测是有缺陷的,这样的预测模型只关注了数据之间深度语义特征,如空间相关性,忽略了序列相关性,而 ARIMA 模型更擅长捕获序列相关性进行短期预测,进而验证了提出的预测框架能够同时利用序列相关性和空间相关性等信息进行建模的有效性。另外,提出的预测框架在 1、3、5 和 7 天长短期 PM2.5 预测和四个指标下的性能始终优于 SVM 模型和 LSTM 模型,产生了最佳的实验结果。

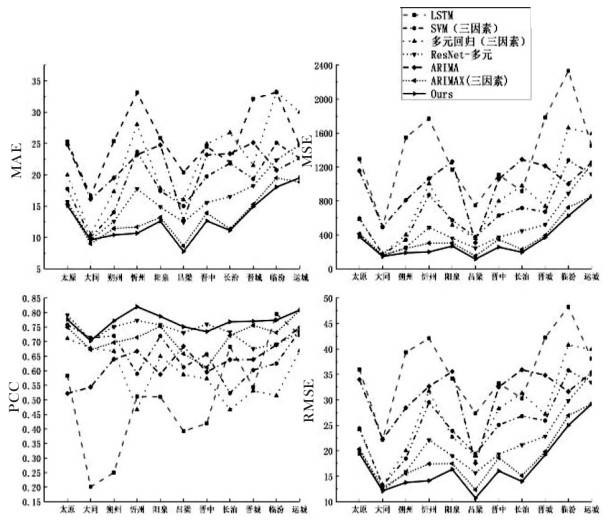


图 5 山西省 11 个城市的七种模型的
四种评价指标比较

最后,给出了预测长度为 1 时,山西省每个城市的七个预测模型 MSE、RMSE、MAE 和 PCC 值的对比结果。如图 5 所示,ARIMA 模型、三因素 ARIMAX 模

型、多元回归模型、ResNet-多元回归模型、LSTM 模型和 SVM 模型在 MSE、RMSE 和 MAE 度量下的线条明显高于提出的预测框架的线条,上述六个模型在 PCC 度量下的线条明显低于提出的预测框架的线条。这进一步证实了提出的预测框架无论是 11 个城市的平均结果还是各个城市独立结果上均展示了最优的性能,进一步验证提出预测框架的有效性。

3 结束语

综上所述,该文提出了一个融合卷积神经网络特征的 ARIMAX 雾霾序列预测框架。运用 ResNet-50 卷积神经网络捕获雾霾影响因素序列中的深度语义特征,如邻近城市的雾霾的空间相关信息,用 ARIMAX 模型建模捕获雾霾的时间序列相关信息。在真实的空气质量和气象数据集(中国山西省)上验证了提出的雾霾 PM_{2.5} 浓度预测框架的有效性。实验结果表明,在 1、3、5、7 天长短期 PM_{2.5} 浓度预测中,提出的预测框架始终优于当前广泛使用的 ARIMA 模型、三因素 ARIMAX 模型、多元回归模型、ResNet-多元回归模型、LSTM 模型和 SVM 模型。

参考文献:

- [1] FENG R, WANG Q, HUANG C C, et al. Ethylene, xylene, toluene and hexane are major contributors of atmospheric ozone in Hangzhou, China, prior to the 2022 Asian Games [J]. *Environmental Chemistry Letters*, 2019, 17: 1151-1160.
- [2] HUANG J, ZHOU C, LEE X, et al. The effects of rapid urbanization on the levels in tropospheric nitrogen dioxide and ozone over East China [J]. *Atmospheric Environment*, 2013, 77(3): 558-567.
- [3] JOYCE T J, GROSSMAN M, GOLDMAN F. An assessment of the benefits of air pollution control: the case of infant health [J]. *Journal of Urban Economics*, 1989, 25(1): 32-51.
- [4] ATKINSON R W, CAREY I M, KENT A J, et al. Long-term exposure to outdoor air pollution and incidence of cardiovascular diseases [J]. *Occupational and Environmental Medicine*, 2015, 72(1): 42-48.
- [5] 蒲国林, 刘笃晋. 基于改进神经网络的环境空气质量预测 [J]. *计算机技术与发展*, 2018, 28(9): 181-184.
- [6] LU K, FUCHS H, HOFZUMAHUS A, et al. Fast photochemistry in wintertime haze: consequences for pollution mitigation strategies [J]. *Environmental Science and Technology*, 2019, 53: 10676-10684.
- [7] GEP G B, BOX G, JENKINS G. Time series analysis: forecasting and control [J]. *Journal of Time*, 1976, 31(4): 238-242.
- [8] 付倩尧. 基于多元线性回归的雾霾预测方法研究 [J]. *计算机科学*, 2016, 43(Z6): 526-528.
- [9] MARTIN R L, OEPPEN J E. The identification of regional forecasting models using space-time correlation functions [J]. *Transactions of the Institute of British Geographers*, 1975, 66: 95-118.
- [10] SIMS C A. Money, income, and causality [J]. *American Economic Review*, 1972, 62(4): 540-552.
- [11] SHAHRIAR S A, KAYES I, HASAN K, et al. Potential of ARIMA-ANN, ARIMA-SVM, DT and CatBoost for atmospheric PM_{2.5} forecasting in Bangladesh [J]. *Atmosphere*, 2021, 12(1): 10.
- [12] 付悦, 夏小娜. 决策分类器在空气质量数据分析中的实证对比 [J]. *计算机技术与发展*, 2020, 30(7): 174-179.
- [13] ZHENG T, BERGIN M H, HU S, et al. Estimating ground-level PM_{2.5} using micro-satellite images by a convolutional neural network and random forest approach [J]. *Atmospheric Environment*, 2020, 230: 117451.
- [14] JIANG T, CHEN B, NIE Z, et al. Estimation of hourly full-coverage PM_{2.5} concentrations at 1-km resolution in China using a two-stage random forest model [J]. *Atmospheric Research*, 2020, 248: 105146.
- [15] HAN W, YANLAN W U, REN F. The prediction of air pollutants based on full connection and LSTM neural network [J]. *Geomatics World*, 2018, 25(3): 34-40.
- [16] 李栋, 张蕾, 郭茂祖, 等. 基于时空卷积残差网络的空气质量预测 [J]. *计算机技术与发展*, 2020, 30(6): 124-129.
- [17] ZHANG B, ZHANG H, ZHAO G, et al. Constructing a PM_{2.5} concentration prediction model by combining auto-encoder with Bi-LSTM neural networks [J]. *Environmental Modelling and Software*, 2020, 124: 104600.
- [18] 朱旭辉, 倪志伟, 倪丽萍, 等. 基于相异度的 SVM 选择性集成雾霾天气预测方法 [J]. *系统科学与数学*, 2017, 37(6): 1480-1493.
- [19] 刘梦炀, 武利娟, 梁慧, 等. 一种高精度 LSTM-FC 大气污染物浓度预测模型 [J]. *计算机科学*, 2021, 48(6A): 184-189.
- [20] LECUN Y, BENGIO Y. Convolutional networks for images, speech, and time series [M]//The handbook of brain theory and neural networks. New York: ACM, 1998: 255-258.
- [21] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]//2016 IEEE conference on computer vision and pattern recognition (CVPR). Las Vegas: IEEE, 2016: 770-778.