DOI:10.19651/j.cnki.emt.2107055

基于机器学习的差异融合分析在空气质量预测中的应用*

高 嵩[1,2] 何卓骏[1] 刘子岳[1] 刘家明[1] 王 刚[1] 李登柯[1]
(1.成都理工大学 机电工程学院 成都 610059; 2.成都理工大学 地球勘探与信息技术教育部重点实验室 成都 610059)

摘 要: 使用机器学习算法对未来 AQI 进行预测,有助于从宏观角度分析未来空气质量变化趋势。在传统上使用单一的机器学习模型对空气质量进行预测时,很难在不同 AQI 波动趋势下都能获得较好的预测效果。为有效解决该问题,在预测方式上进行改进,针对使用随机森林模型和基于卷积神经网络和注意力机制的长短期记忆模型对成都市的 AQI 数据进行预测时,在不同的 AQI 波动趋势下两者的预测准确度不同的特点,设计了一种差异融合分析模型。实验结果表明,提出的差异融合分析模型的 MSE 误差较随机森林模型降低了 5.8%,较基于卷积神经网络和注意力机制的长短期记忆模型降低了 6.3%。

关键词: 空气质量指数;差异融合;随机森林;长短期记忆模型;支持向量机
中图分类号: X51;TP391 **文献标识码**: A **国家标准学科分类代码**: 610.30

Application of difference fusion analysis based on machine learning in air quality prediction

Gao Song[1,2] He Zhuojun[1] Liu Ziyue[1] Liu Jiaming[1] Wang Gang[1] Li Dengke[1]
(1. School of Mechanical and Electrical Engineering, Chengdu University of Technology, Chengdu 610059, China;
2. Key Lab of Earth Exploration & Information Techniques of Ministry of Education, Chengdu University of Technology, Chengdu 610059, China)

Abstract: The prediction of AQI in the future by using machine learning algorithm is helpful to analyze the trend of air quality change in the future from a macro perspective. When a single machine learning model is traditionally used to predict air quality, it is difficult to obtain good prediction results under different AQI fluctuation trends. In order to effectively solve this problem, the prediction method is improved. When using random forest model and long and short-term memory model based on convolution neural network and attention mechanism to predict the AQI data in Chengdu, a difference fusion analysis model is designed according to the characteristics of different prediction accuracy under different AQI fluctuation trends. The experimental results show that the MSE of the proposed difference fusion analysis model is 5.8% lower than that of the random forest model, and 6.3% lower than that of the long-term and short-term memory model based on convolutional neural network and attention mechanism.

Keywords: air quality index;difference fusion;random forest;long short-term memory;support vector machines

0 引 言

随着中国城镇化和工业化的快速发展,中国经济飞速发展,但随之而来的环境问题也日益严峻[1]。在全球气候急剧变化的大背景下,大部分区域地表风速减小,空气质量多次触及红线,城市居民的易居水平急剧下降,严重损害人民群众的身体健康[2]。

空气质量指数(air quality index, AQI)是定量描述空气质量状况的无量纲指标,它能够直观评价大气的环境污染水平。AQI 也是一种描述空气中污染物浓度的定量化数值,通常情况下,AQI 的数值越大,空气中污染物浓度就越高,对健康的伤害就越大,人体舒适程度就越低[3-4]。

成都市位于我国西南部的四川盆地。四川盆地具有特殊的地势,西部紧邻青藏高原东部边缘,使得四川盆地风速

收稿日期:2021-06-24
* 基金项目:国家自然科学基金(41930112)项目资助

较低、湿度较大，每当秋冬季时，“上暖下冷”的逆温结构尤为明显，致使四川盆地的大气污染程度以及大气污染日数易在秋冬季节增加，因而对位于四川盆地的成都市的空气质量进行预测、预估十分重要[5-6]。

目前学界已将机器学习应用于空气质量的预测。中浩洋等[7]在 2014 年使用 BP 人工神经网络(BPNN)对环境污染物 SO_2 浓度进行预测，预测结果能较好地反映 SO_2 浓度的变化规律，但当污染物浓度突发性升高时的预测能力较弱。谢永华等[8]在 2015 年利用支持向量机对城市的 PM2.5 浓度进行预测，相对于神经网络(NN)等其他机器学习模型，支持向量机回归(SVR)方法在中短期的预测应用上有着较好的表现，但在特征维度较高的情况下，模型较为复杂。杨瑞君等[9]在 2017 年基于随机森林模型对城市空气质量进行评价，建立了空气质量评价因子与空气质量等级之间的内在映射关系，评价预测的准确性较高。石晓文等[10]在 2019 年从空气质量等级维度分析，得出长短时记忆网络(long short-term memory，LSTM)空气质量预测模型相对于 BPNN 模型、SVR 模型等来说，在空气质量指数较低时的预测效果更好，并且提出了可以根据不同的空气质量等级选择不同的预测算法进行预测的思想。袁燕等[11]在 2020 年提出一种基于社区划分的空气质量指数预测的算法，该算法提高的预测精度，降低了时间复杂度。Su 等[12]在 2020 年建立了基于遗传算法和 BP 神经网络的空气质量指数预测模型，对 AQI 的预测研究具有一定的指导意义。目前，对于空气质量的预测，学术界使用的主流机器学习模型有神经网络、支持向量机回归、随机森林等，这些模型能够对空气质量进行较为精准的预测。但是，各个模型在同一变化趋势下的预测准确度不同，特别是空气质量数据在某一个时间范围内发生突发性改变的时候，模型的预测效果差异很大。

本文首先制作了成都市 2016 年～2021 年 AQI 数据集，分别使用随机森林(random forest，RF)模型和基于 CNN 与 LSTM-Attention 的混合机制(CNN-LSTM attention，CLA)模型对成都市空气质量进行预测，两种模型的预测效果均较好，对比发现：AQI 数值波动较大时，CLA 预测能力相对较弱；AQI 数值波动较小时，RF 预测能力相对较弱。对此，本文提出了一种差异融合分析(fusion difference analysis，DFA)模型。该模型通过使用融合阈值搜寻算法寻找的最佳融合阈值，对 RF 模型的预测结果和 CLA 模型的预测结果进行融合，融合后的预测结果对于任何 AQI 数值波动都具有更好的表现，采用 MAE、MSE、RMSE、R2 四个评价指标对比评价了 DFA 模型、CLA 模型、RF 模型和 SVM 模型的预测结果，评价结果表明使用 DFA 模型预测成都市 AQI 具有更高的精度。

1　成都市 AQI 数据集的制作

我国《环境空气质量指数(AQI)技术规定(试行)》(HJ 633—2012)对空气质量分指数(IAQI)及对应的污染物项目浓度限值均有规定，本文使用的主要参数如表 1 所示，其中 IAQI 无单位，污染物浓度限值单位均为 $\mu g \cdot m^{-3}$[13]。

表 1　空气质量分指数及对应的污染物项目浓度限值

IAQI	CO (24 h 平均)	NO_2 (24 h 平均)	O_3 (1 h 平均)	PM_{10} (24 h 平均)	$PM_{2.5}$ (24 h 平均)	SO_2 (24 h 平均)
0	0	0	0	0	0	0
50	2 000	40	160	50	35	50
100	4 000	80	200	150	75	150
150	14 000	180	300	250	115	475
200	24 000	280	400	350	150	800
300	36 000	565	800	420	250	1 600
400	48 000	750	1 000	500	350	2 100
500	60 000	940	1 200	600	500	2 620

AQI 的计算采用式(1)，n 代表污染物项目。

$$AQI = \max\{IAQI_1, IAQI_2, IAQI_3, \cdots, IAQI_n\} \quad (1)$$

为了训练和测试评价各模型对成都市 AQI 的预测性能，需要制作成都市 AQI 数据集。成都市 AQI 数据来源于中华人民共和国生态环境部，选取 2016 年 1 月 1 日～2020 年 3 月 31 日共计 1 552 天(实为 1 546 天)数据作为训练集，其中缺少的单日 AQI 数据 X_n 采用前后两日 AQI 数据 X_{n-1} 和 X_{n+1} 的平均值作为该日的 AQI 数据，其计算如式(2)所示。

$$X_n = \frac{X_{n-1} + X_{n+1}}{2} \quad (2)$$

缺少的连续两日 AQI 数据 X_n 和 X_{n+1}，采用前后 4 日 AQI 数据 X_{n-1}、X_{n-2} 和 X_{n+2}、X_{n+3} 的平均值计算这两日的 AQI 数据，其计算式如式(3)～(5)所示。

$$AVG = \frac{\sum_{i=n-2}^{n-1} X_i + \sum_{i=n+2}^{n+3} X_i}{4} \quad (3)$$

$$X_n = \frac{AVG + X_{n+1}}{2} \quad (4)$$

$$X_{n\,1}=\frac{AVG+X_{n\,12}}{2} \tag{5}$$

按上述方法将数据补充完整,如图 1 所示,展示了训练集中 AQI 和各项污染物浓度值的变化情况。

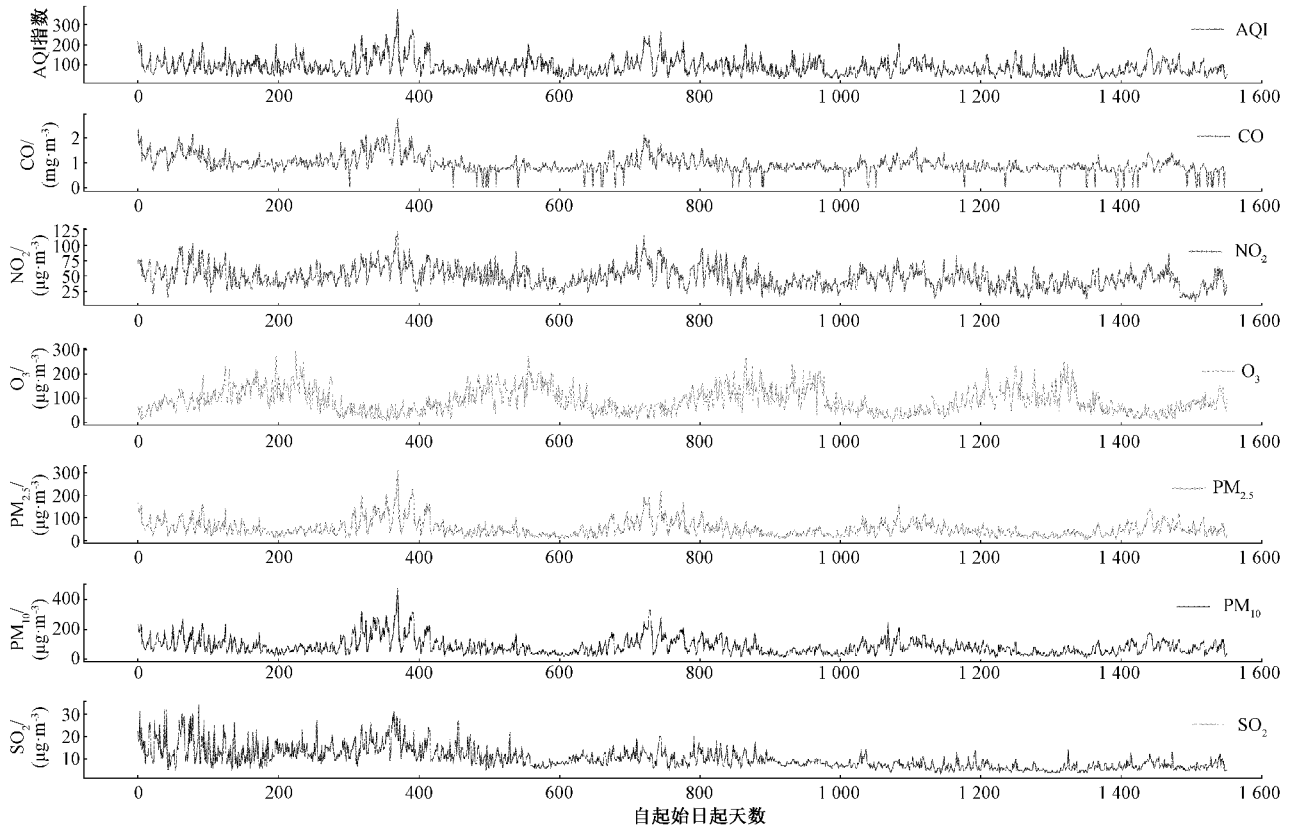


图 1　成都市 AQI 训练集中 AQI 和各项污染物浓度值的变化情况

将训练集按照 20∶1 的比例分离出少量的数据作为验证集。另外,获取 2020 年 4 月 15 日～2021 年 3 月 31 日共 351 天的成都市 AQI 和各项污染物浓度值作为测试集,用于评价各个模型的预测性能。因此,本文所用的成都市 AQI 数据集(CD-AQI)由 1 552 天训练集(含验证集)和 351 天测试集组成。

另外,在 CLA 模型进行数据调用前,需要对数据集进行归一化处理,使数据落入一个小的特定区间 0～1 范围之间,如式 6 所示。

$$\tilde{X}=\frac{(X-X_{\min})}{X_{\max}-X_{\min}} \tag{6}$$

式中:X 代表每一个数据量,$X_{\min}$ 代表数据集中的最小数据值,$X_{\max}$ 代表数据集中的最大数据值,$\tilde{X}$ 代表输出的归一化之后的数据值。

2　RF 模型和 CLA 模型的 AQI 预测

2.1　RF 模型和 CLA 模型

1)RF 模型

RF 由 Leo Breiman 提出,是 Bagging 算法的一个拓展,可以解决多类分类的问题,运用了集成学习的思想,即将多个弱分类器结合成为一个强分类器[14-15]。通过整合模型的各种参数,解决了模型代表单一的问题,避免了判断时出现的局限性[16-17]。

以 MSE 均方差作为衡量指标,通过随机搜索确定 MSE 最小时 RF 模型的最佳参数。随机搜索表示以随机的方式在参数空间中采样搜索;对于连续变量的参数,随机搜索会进行分布式采样,采样完成后进行交叉验证(cross validation, CV),通过比较每一种设置的参数下的训练器精度,最终选择最优的参数值。其中树的最大深度设置为 10,允许分枝时一个节点必须包含的最小训练样本数值设置为 10,分枝后的子节点的训练样本数的最低数目设置为 1。

RF 模型输入由前一天的 7 项空气质量指标(AQI,CO,NO_2,O_3,PM_{10},$PM_{2.5}$,SO_2)组成,输出标签是后一天的 AQI 值。采用搜寻到的最佳参数,在训练集上完成模型的训练。

2)CLA 模型

CLA 模型主要由 CNN 模块、LSTM 模块以及 ATTENTION 模块构成。

根据上述 3 个模块对 CLA 模型进行结构搭建,首先将输入的数据传入 CNN 网络中,对训练数据进行特征提取再传入 LSTM 网络进行进一步处理。数据传入 LSTM 层

后，LSTM 层会保留传入数据的有用信息，遗忘无用信息，并将处理后的数据传入 ATTENTION 层，对数据中的关键信息给予足够的关注，突出关键信息的影响，提高模型预测准确性[18-20]。最后经过全连接层后输出预测结果。其中 Dropout 作用是防止模型训练过拟合。

本文使用的 CLA 模型网络结构如图 2 所示。

利用 CD-AQI 的训练集对 CLA 模型进行训练。模型的输入由前 14 天的 7 项空气质量指标（AQI，CO，NO_2，O_3，PM_{10}，$PM_{2.5}$，SO_2）组成，输出标签是后一天的 AQI 值。训练时设置损失函数为 MSE，优化器采用自适应矩估计优化算法（Adam 算法），该方法具有更快的收敛速度和更低的内存消耗。最终获得并保存训练完成的 CLA 模型。

2.2　RF 模型与 CLA 模型的预测结果对比

将训练好的 RF 模型和 CLA 模型应用测试集进行 AQI 预测。如图 3 所示，3 条折线分别表示实际 AQI 数据、CLA 模型预测数据和 RF 模型预测数据。对测试集数据进行遍历，基于 MSE 评价指标评价预测误差，若某一天采用 RF 模型得到的 AQI 预测误差小于 CLA 模型的 AQI 预测误差，则将当天标注成灰色柱线。显然，灰色柱线表示这一天 RF 模型的 AQI 预测误差小，白色柱线表示这一天 CLA 模型的 AQI 预测误差小。

从图 3 中可以看出，灰色柱线密集区的实际 AQI 数据变化较大，采用 RF 模型具有相对较好的预测性能，而灰色柱线稀疏区（即白色柱线密集区）的实际 AQI 数据变化较小，采用 CLA 模型具有相对较好的预测性能。鉴于此，如果将两种模型预测误差相对较小的结果取出，即搜寻合适的阈值对两个模型的预测结果进行融合，就可以获得比原来两个模型更好的预测性能。为此，本文基于预测结果的差异设计了一种 DFA 模型以提高成都市 AQI 预测精度。

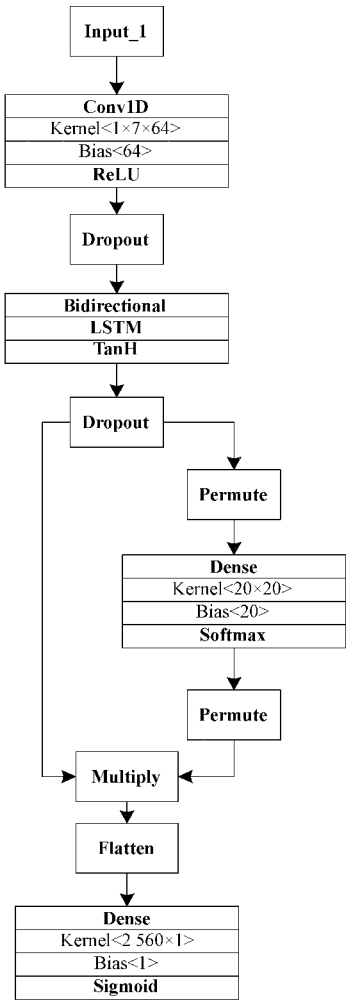


图 2　CLA 模型的网络结构

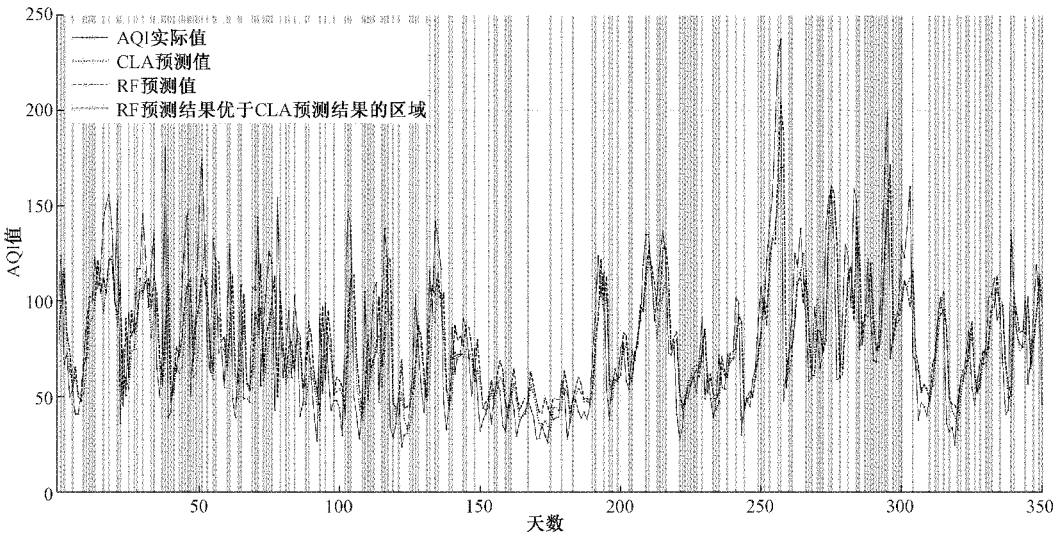


图 3　两种模型的 AQI 预测值以及预测误差的比对

3　DFA 预测方法

基于 RF 模型以及 CLA 模型预测结果并采用差异融

合法的模型称为 DFA 模型,包括阈值搜寻算法(fusion threshold search)与差异融合算法(differential fusion)。通过 RF 和 CLA 模型在 ΔAQI 数值波动程度不同时预测偏差不同的特点,经过阈值搜寻算法计算出融合阈值,再对模型的预测结果使用差异融合算法进行融合,从而使得融合后输出的 AQI 预测值偏差更小,准确度更高,并将融合后输出的预测结果作为模型的最终预测结果。差异融合模型流程如图 4 所示。

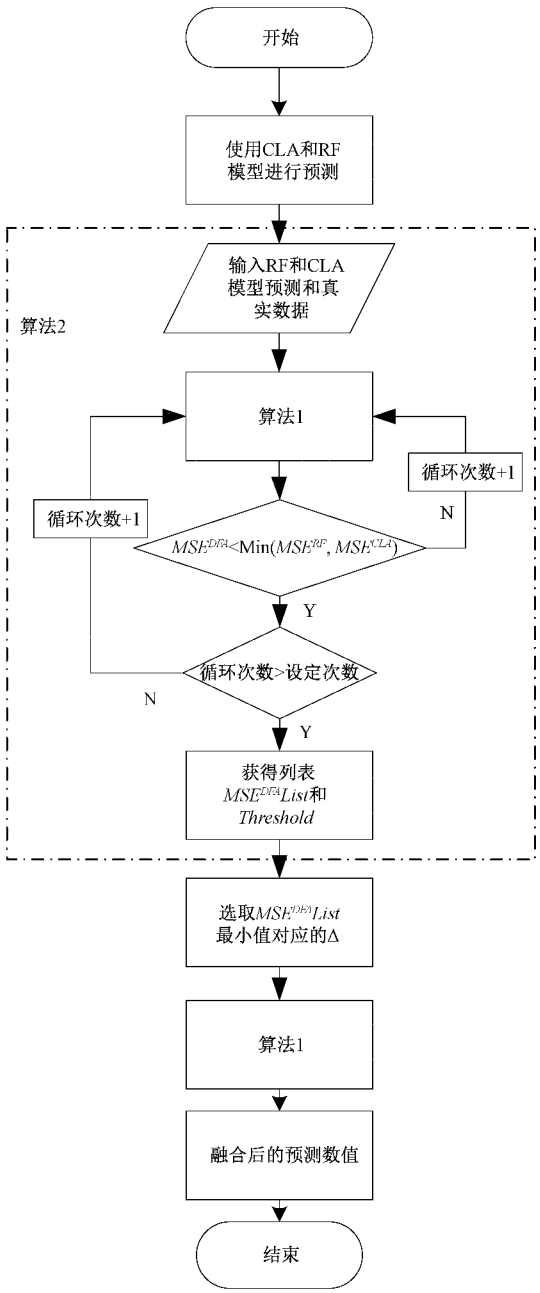


图 4　差异融合模型

DFA 模型通过阈值搜寻算法搜寻并获得一个合适的 AQI 波动阈值,利用这个阈值对 RF 模型和 CLA 模型的预测结果进行融合,作为整个 DFA 模型的最终预测输出。

3.1　差异融合算法

RF 模型和 CLA 模型通过训练,都可以对 AQI 进行较为准确的预测。但当 ΔAQI 的数值在一定范围内波动程度不同时,RF 模型和 CLA 模型的预测效果不同。通过对上述两种模型的预测结果进行 MSE 的计算,比较得出 CLA 模型在 AQI 数值波动较小,且低于某个波动阈值时预测效果要优于 RF 模型的预测效果,RF 模型在 AQI 数值较大,且高于某个波动阈值时预测结果优于 CLA 模型的预测结果。基于 MSE 评价指标,对比使用预测天数前 2 天的 AQI 的差值、前 3 天的 AQI 的平均差值和前 3 天中 AQI 最大值与最小值的差值等差值选取方式,最终选取了 MSE 表现最好的指标:预测天数前 2 天的 AQI 差值,作为本文的波动大小表示方式。为了得到融合后的数值,本文提出了差异融合算法,其中关键步骤如算法 1 所示。

算法 1　差异融合

INPUT:

RF prediction: $B^{RF}=\{b_1^{RF},b_2^{RF},\cdots,b_n^{RF}\}$

CNN-LSTM ATTENTION: $B^{CLA}=\{b_1^{CLA},b_2^{CLA},\cdots,b_n^{CLA}\}$

Real data: $A=\{a_1,a_2,\cdots,a_n\}$

ΔAQI change value: Δ

OUTPUT:

1: Fusion Result: R
2: $R\leftarrow 0$
3: **for** $b_i^{RF}\in B^{RF}$, $b_i^{CLA}\in B^{CLA}$ **and** $a_i\in A$ **do**
4:　**if** $|a_{i-1}-a_{i-2}|<\Delta$ **then**
5:　　$R\leftarrow R+b_i^{CLA}$
6:　**else**
7:　　$R\leftarrow R+b_i^{RF}$
8: **return** R

B^{RF} 代表 RF 模型的预测值集合,$b_n{}^{RF}$ 代表 RF 模型的每一个预测值。B^{CLA} 代表 CLA 模型的预测值集合,$b_n{}^{CLA}$ 代表 CLA 模型的每一个预测值,A 代表真实值的集合,a_n 代表每一个真实值的大小,Δ 表示 AQI 的波动差值大小的阈值。在 Δ 设置完毕后,依次遍历集合 A 中的每一个元素 a_n,计算前两天的 AQI 差值($a_{n\ 1}-a_{n\ 2}$),并取绝对值,当差值小于所设阈值 Δ 时选用 CLA 模型的预测结果,当差值大于所设阈值 Δ 时选用 RF 模型的预测结果,通过循环形成最终的预测结果。

3.2　融合阈值搜寻算法

差异融合算法中最重要的部分是融合阈值的选取,在算法 1 的基础上通过改变阈值 Δ 来寻找最佳的融合阈值,从而使融合后的偏差达到最低。融合阈值搜寻算法如算法 2 所示。

算法 2　融合阈值搜寻

INPUT:

RF prediction: $B^{RF}=\{b_1^{RF},b_2^{RF},\cdots,b_n^{RF}\}$

CNN-LSTM ATTENTION: $B^{CLA}=\{b_1^{CLA},b_2^{CLA},\cdots,b_n^{CLA}\}$

Fusion prediction: $B^{DFA}=\{b_1^{DFA},b_2^{DFA},\cdots,b_n^{DFA}\}$

Real data: $A=\{a_1,a_2,\cdots,a_n\}$

AQI change value: Δ

OUTPUT:

1: $MSE^{DFA}List \leftarrow 0$
2: $Threshold \leftarrow 0$
3: **for** i in range (0,40,1) **do**
4:　**Differential Fusion** (A,B^{CLA},B^{RF},i)
5:　**if** $MSE^{DFA} < \mathrm{Min}(MSE^{RF},MSE^{CLA})$ **then**
6:　　$MSE^{DFA}List \leftarrow MSE^{DFA}List + MSE^{DFA}$
7:　　$\Delta \leftarrow i$
8:　　$Threshold \leftarrow Threshold + \Delta$
9: **return** $MSE^{DFA}List, Threshold$

首先建立两个列表 MSE$^{\mathrm{DFA}}$List 和 Threshold 分别来储存 DFA 模型的 MSE 值和阈值 Δ 的大小。设置循环变量 i(i 从 0～40 以 1 为步长变化),在每次循环中运行算法 1 的融合算法,得到 DFA 模型的预测值,并分别计算 DFA、CLA、RF 模型的 MSE 值。当 DFA 模型的 MSE 值小于 RF 和 CLA 模型的 MSE 值的最小值时,将此时的 DFA 模型的 MSE 值和对应的阈值 Δ 分别存入列表 MSE$^{\mathrm{DFA}}$List 和 Threshold 中,直到循环结束。

在获得 MSE$^{\mathrm{DFA}}$List 和 Threshold 列表的最终值后,选取里面的 MSE 的最小值,和对应的阈值 Δ,则阈值 Δ 为最终需要的融合阈值。

4　预测试验及性能评价

4.1　预测性能评价方法及指标

本文研究的目的是利用往日的空气质量数据,预测未来一天的 AQI 数值,即预测明天的空气质量。为了衡量 DFA 模型的 AQI 预测性能,采用绝对平均误差(mean absolute error, MAE)、均方误差(mean squared error, MSE)、均方根误差(root mean squared error, RMSE)以及决定系数 R2 作为评价指标,其计算公式分别如式(7)～(10)所示。其中,y_i 为第 i 天 AQI 实际数值,$\hat{y}_i$ 为第 i 天 AQI 预测值,$\bar{y}_i$ 是 AQI 实际数据的平均值。

MAE、MSE、RMSE 数值越小,R2 数值大小越接近 1,表明模型的预测精度越高。

$$MAE=\frac{1}{N}\sum_{i=1}^{N}|y_i-\hat{y}_i| \tag{7}$$

$$MSE=\frac{1}{N}\sum_{i=1}^{N}(y_i-\hat{y}_i)^2 \tag{8}$$

$$RMSE=\sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i-\hat{y}_i)^2} \tag{9}$$

$$R^2=1-\frac{\sum_i(y_i-\hat{y}_i)^2}{\sum_i(y_i-\bar{y}_i)^2} \tag{10}$$

4.2　预测实验

前文已知 CD-AQI 数据变化波动较大,采用本文提出的 DFA 模型预测性能更好,为了证明所使用数据集的非偶然性,选取成都市 2015 年全年的 AQI 数据作为融合阈值搜寻算法的输入,运行该算法,得到 DFA 模型的最佳融合阈值大小为 20。

为了验证 DFA 模型具备更好的预测性能,本文使用 RF、CLA、DFA 模型以及支持向量机(support vector machines, SVM)模型[21-22]对 CD-AQI 进行训练和测试。如图 5 所示,从上到下分别是 RF、SVM、CLA、DFA 4 种模型对 CD-AQI 测试集(351 天)的预测结果,实线表示实际 AQI 数据,虚线表示不同模型的 AQI 预测结果。

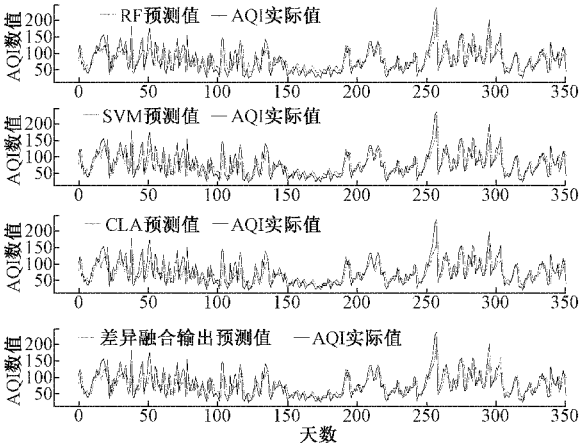
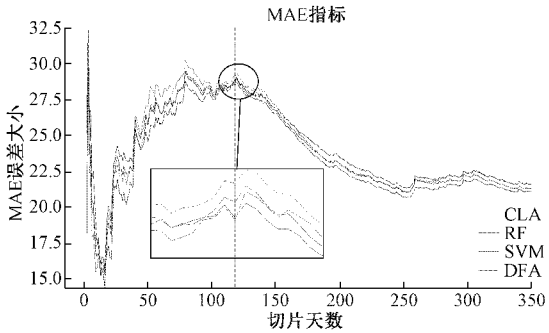


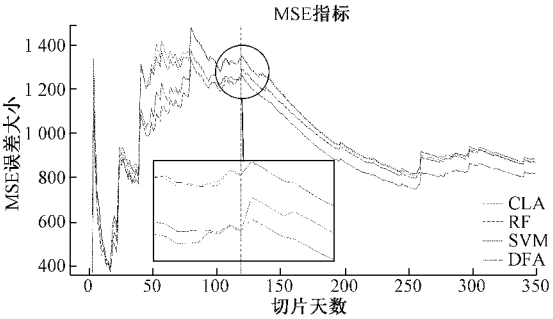
图 5　RF、SVM、CLA、DFA 4 种模型的 CD-AQI 测试集预测结果

进一步对 RF、CLA、SVM、DFA 4 种模型的预测结果进行预测性能评价,通过时间切片,即从 2021 年 4 月 15 日开始,依次对 351 天截取连续时间的数据,得到最小切片 1(2 天,2020 年 4 月 15 日～4 月 16 日),切片 2(3 天,2020 年 4 月 15 日～4 月 17 日),…,最大切片 350(351 天,2020 年 4 月 15 日～2021 年 3 月 31 日),分别计算 4 种模型在 350 个不同长度时间切片的 MAE、MSE、RMSE、R2 指标表现,其结果如图 6(a)～(d)所示,横坐标是切片天数,范围为 2～351 天,纵坐标代表某一种评价指标数值。

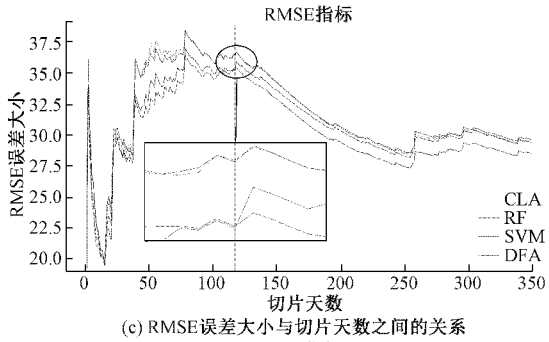
如图 6(a)～(d)所示,当 DFA 模型的 MAE、MSE、RMSE、R2 评价指标劣于 RF 模型、CLA 模型和 SVM 模型时,表示此切片天数以内的融合精度相对较差的时间切片个数占比大于 50%;反之,当 DFA 模型的 MAE、MSE、



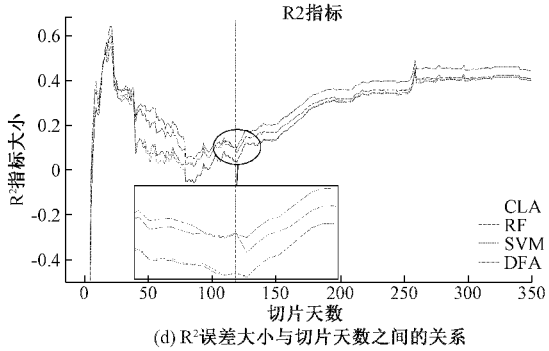
(a) MAE误差大小与切片天数之间的关系



(b) MSE误差大小与切片天数之间的关系



(c) RMSE误差大小与切片天数之间的关系



(d) R^2误差大小与切片天数之间的关系

图 6　4 种评价指标与切片天数大小的关系

RMSE 评价指标优于 RF 模型、CLA 模型和 SVM 模型时，表示此切片天数以内的融合精度相对较好的时间切片个数占比大于 50%。

从图 6 中可以看出，当切片天数小于 118 天时，DFA 模型 4 项指标有时候优于其他模型，有时候劣于其他模型，预测性能不稳定；切片天数从第 118 天开始，DFA 模型的 4 项指标均优于其他模型，预测性能非常稳定、可靠。可见，样本数量到达一定的时候，DFA 模型的综合表现优于其他 3 种模型。对比测试集中一年中不同模型的数据指标，DFA 模型的 MSE 误差较 RF 模型降低了 5.8%，较 CLA 模型降低了 6.3%。因此，说明了本文提出的 DFA 模型总体上能够更好地预测成都市空气质量，能够相对于原来的模型更好地反映成都市未来的 AQI 变化。

5　结　　论

本文运用 RF 模型和 CLA 模型对成都市 AQI 预测，由于 AQI 数据波动较大导致两种模型预测结果呈现差异化表现，基于该特点提出了一种差异融合模型，通过多种模型对成都市 AQI 数据集测试，结果表明本文提出的模型预测性能更好。

制作了成都市 AQI 数据集。数据集中包括了成都市 2016 年 1 月 1 日～2021 年 3 月 31 日这段时间内 7 项空气质量指标(AQI，CO，NO_2，O_3，PM_{10}，$PM_{2.5}$，SO_2)数据。

运用 RF 模型和 CLA 模型对成都市 AQI 预测，对比结果可看出 RF 模型在 AQI 较大波动时的预测较为准确，误差较低，拟合程度较好；CLA 模型在 AQI 较小波动时的预测较为准确，误差较低，拟合程度较好。基于该差异化特点，提出了基于 RF 模型和 CLA 模型结果构造的使用差异融合法的 DFA 模型，给出了具体算法。

使用 4 种模型对数据集测试，采用 MAE、MSE、RMSE、R2 评价指标，通过时间切片计算了 4 种模型不同时间段的指标。在样本数目足够多时，分析出 DFA 模型的预测性能更优，且非常稳定、可靠。对于成都市空气质量指数预测中有很好的实用价值。

由于差异融合分析模型的特点，本文方法的使用只局限于 AQI 波动变化比较显著的城市。

参考文献

[1]　程浩. 中国城乡人居环境：成绩与挑战[J]. 人类居住，2016(3)：22-23.

[2]　王帅，杜丽，王瑞斌，等. 国内外环境空气质量指数分析和比较[J]. 中国环境监测，2013，29(6)：58-65.

[3]　高庆先，刘俊蓉，李文涛，等. 中美空气质量指数(AQI)对比研究及启示[J]. 环境科学，2015，36(4)：1141-1147.

[4]　丁玉贤，孙维娜，牛建刚. 2014 年～2019 年呼包鄂城市群 AQI 表征研究[J]. 内蒙古科技与经济，2020(14)：7-8.

[5]　周子涵，张小玲，康平，等. 基于异常天气分析法探究四川盆地冬季大气污染的气象成因[J]. 安全与环境工程，2020，27(2)：66-77.

[6]　张莹，王式功，倪长健，等. 成都冬季 $PM_{2.5}$ 污染天气形势的客观分型研究[J]. 环境科学与技术，2020，43(5)：139-144.

[7]　申浩洋，韦安磊，王小文，等. BP 人工神经网络在环境空气 SO_2 质量浓度预测中的应用[J]. 环境工程，

2014,32(6):117-121.
[8] 谢永华,张鸣敏,杨乐,等.基于支持向量机回归的城市 $PM_{2.5}$ 浓度预测[J].计算机工程与设计,2015,36(11):3106-3111.
[9] 杨瑞君,赵楠,凡耀峰,等.基于随机森林模型的城市空气质量评价[J].计算机工程与设计,2017,38(11):3151-3156.
[10] 石晓文,蒋洪迅.面向高精度与强鲁棒的空气质量预测 LSTM 模型研究[J].统计与决策,2019,35(16):49-53.
[11] 袁燕,陈伯伦,朱国畅,等.基于社区划分的空气质量指数(AQI)预测算法[J].南京大学学报(自然科学版),2020,56(1):142-150.
[12] SU Y, XIE H. Prediction of AQI by BP neural network based on genetic algorithm[C]. 2020 5th International Conference on Automation, Control and Robotics Engineering(CACRE), 2020: 625-629.
[13] 潘本锋,李莉娜.环境空气质量指数计算方法与分级方案比较[J].中国环境监测,2016,32(1):13-17.
[14] 徐肖伟,李鹤健,于虹,等.基于随机森林的变压器油中溶解气体浓度预测[J].电子测量技术,2020,43(3):66-70.
[15] 杨傲雷,刘佳奇,徐昱琳,等.融合随机森林模型的单目视觉人体空间定位方法[J].仪器仪表学报,2020,41(11):207-215.
[16] 杨思琪,赵丽华.随机森林算法在城市空气质量预测中的应用[J].统计与决策,2017(20):83-86.
[17] 程蓉,钱雪忠.基于神经随机森林的局部空气质量预测模型[J].计算机工程与设计,2020,41(7):1958-1966.
[18] CAI R, ZHU B, JI L, et al. An CNN-LSTM attention approach to understanding user query intent from online health communities[C]. 2017 IEEE International Conference on Data Mining Workshops (ICDMW), 2017: 430-437.
[19] 张春露.基于 Tensorflow 的 LSTM 在太原空气质量 AQI 指数中的分析与预测[D].太原:中北大学,2019.
[20] 魏昱洲,许西宁.基于 LSTM 长短期记忆网络的超短期风速预测[J].电子测量与仪器学报,2019,33(2):64-71.
[21] 沈舒,吴聪,李勃,等.基于优化 SVM 的高速公路交通事件检测[J].电子测量技术,2012,35(5):40-44,82.
[22] 李帅,王瑛,薛占龙,等.基于 LS-SVM 的接地电阻监测数据回归预测方法[J].国外电子测量技术,2019,38(8):19-22.

作者简介

高嵩,工学博士,副教授,主要研究方向为无人机目标检测及识别等。
E-mail:gs@cdut.edu.cn

何卓骏,硕士研究生,主要研究方向为无人机目标检测及识别等。
E-mail:644416883@qq.com