doi:10. 16055/j. issn. 1672-058X. 2021. 0006. 016

基于随机森林回归和气象参数的城市空气质量预测模型
——以重庆市为例

徐艳平, 陈义安
(重庆工商大学 数学与统计学院,重庆 400067)

摘 要:为有效进行城市空气质量预测、推进城市空气污染防治, 弥补传统统计学模型在大数据时代背景下对城市空气质量预测准确率低、容错能力差等问题,提出利用随机森林回归构建城市空气质量预测模型;综合考量污染物浓度、气象参数、时间参数等多方面影响因素,通过网格搜索法调整参数的最优组合,构建基于随机森林回归算法的城市空气质量预测模型;基于重庆市 2017-01-01—2020-07-31 的指标数据,对重庆市空气质量进行预测分析,结果表明:在模型下训练集与测试集的确定性系数 R^2 均在 99% 以上,均方误差 D_{MSE} 和平均绝对误差 D_{MAE} 在训练集和测试集上的取值均在可接受范围内,证实模型具有运行速度快、预测误差小、具有较高的预测精度等优点,具备较好的学习能力与泛化能力。

关键词:随机森林回归;空气质量预测;气象参数;空气质量指数
中图分类号:X831 **文献标志码**:A **文章编号**:1672-058X(2021)06-0118-07

0 引 言

伴随工业的发展、化石燃料的消费以及城市化进程的不断加快,中国城市空气污染问题日趋严重,给人们的生活生产带来极坏的影响。探索高效率高准确率的城市空气质量预测模型对推进城市空气污染防治意义重大。

目前,城市空气质量的预测模型主要分为三大类:通过总结前人的研究经验结合大气运动等进行预测、基于传统的统计学模型预测和基于机器学习等深度学习算法预测。沈劲等[1]将聚类方法与多元回归方法相结合进行空气质量预测,发现其具有较好的预测准确率;李博群等[2]利用模糊时间序列预测南京市的空气质量指数,并证实了该模型的预测准确率;汤泽梅[3]为构建空气质量预测模型,选取乌鲁木齐地区的空气质量指数为因变量,使用多元线性回归方法分析了相关指标数据;王建书等[4]运用 ARIMA 模型预测苏州市的空气质量指数并取得了较为理想的预测结果。上述学者均采用基于传统统计学方法构建的城市空气质量预测模型,这些方法在大数据时代背景下,模型容错能力较低、预测准确率相对较差,无法满足对海量数据进行智能化处理的需求。

伴随当下人工智能技术的不断发展,大数据时代已悄然来临,面对海量气象数据,更多学者选择使用机器学习方法构建城市空气质量预测模型。司志娟[5]构建了基于因素分析的灰色神经网络组合模型来预测空气质量;赵李明[6]将遗传算法与 BP 神经网络相结合,用于研究广州市空气质量预测并取

收稿日期:2020-10-17;**修回日期**:2020-12-10.
作者简介:徐艳平(1995—),女,重庆巫山人,硕士研究生,从事经济统计研究.

得了较为准确的结果;张楠等[7]改进了灰狼算法并将其与支持向量机相结合,用于进行城市空气质量预测模型的构建;夏润等[8]通过改进 XGBoost 算法构建了预测能力与泛化性都比较优异的空气质量预测模型;郑洋洋[9]构建了基于深度学习库 Keras 的长短期记忆循环神经网络预测模型,并较准确地预测了太原市空气质量指数;徐旭冉[10]运用决策树算法构建了将所有污染参数作为评估空气质量因素的空气质量预测模型。基于机器学习等深度学习算法构建的城市空气质量预测模型有着预测准确率高、数据处理能力强等优点,成为当下空气质量预测模型构建的主要方法。

随机森林算法作为一种取代神经网络等传统机器学习方法的分类回归算法,具有高准确率、不易过度拟合、对噪声及异常值容忍度高等特点。相比于传统的多元线性回归模型,随机森林算法能够克服协变量之间复杂的交互作用,且勿需预先设定函数形式;相比于神经网络,随机森林算法不易过度拟合;相比于支持向量机,随机森林算法规避了支持向量机核函数及内部函数依赖于使用者技巧的问题,因此随机森林算法被广泛应用于各领域研究并取得较好效果。孔丽英等[11]基于企业进销项发票数据,采用随机森林算法构建了税收风险预测模型;Sanjiban Sekhar Roy 等[12]分别运用随机森林算法、梯度提升机和深度神经网络进行股票价格进行预测;李刚在研究电力负荷预测时对随机森林的决策树进行了基于遗传算法的改进,从而大幅度降低预测时间消耗;Koutarou Matsumoto[13]‖运用随机森林算法进行了急性缺血性卒中后脑卒中预后评分和临床结果的数据驱动预测;马冉等[14]利用随机森林算法对三峡库区草堂河流域土壤的 pH 值进行了空间分布预测,结果显示其平均绝对误差低、预测精度高,能够作为预测土壤 pH 值的有效方法。

基于此,选择区别于传统统计学方法与传统机器学习方法的随机森林算法构建城市空气质量预测模型,并相较于传统模型仅考虑大气污染物浓度,选择时间参数、气象参数及大气污染物浓度为城市空气质量预测模型影响因素,有效提升预测准确率和数据处理效率,为空气污染的防控治理提供更为准确的空气质量信息。

1　随机森林回归算法

随机森林(Random Forest,RF)算法是通过构建多棵决策树形成森林的一种分类与回归算法。它以决策树为基本单元,选取 bootstrap 重采样方法随机得到多个互不相同的样本子集,采用随机子空间划分的方法依据各样本子集构建决策树。构建决策树时的特征由全部特征随机抽取得到,当决策树进行节点分裂时,选取随机生成的特征子集中的最优特征进行分裂。最后对所有决策树的预测结果采取众数投票或者取平均值,得到随机森林最终的预测结果。简单来讲,随机森林就是由多个弱学习器(决策树)所集成的强学习器。

设随机向量(X,Y)是独立分布的。从(X,Y)中随机生成训练集,输入向量与输出向量分别为 X、Y,则预测结果 $h(X)$的均方泛化误差表示为

$$E_{X,Y}[Y-h(X)]^2$$

随机森林回归的预测结果是 k 棵决策树的预测结果$\{h(\theta,X_k)\}$取均值而来的,它满足以下定理:

定理 1　当 $k\to\infty$,

$$E_{X,Y}[Y-\bar{h}_k(X,\theta_k)]^2\to E_{X,Y}[Y-E_\theta(X,\theta_k)]^2 \quad (1)$$

式(1)右侧部分表示随机森林的泛化误差,将其记为 PE^{**}。PE^* 则表示一棵决策树的平均泛化误差,即

$$PE^*=E_\theta E_{X,Y}[Y-h(X,\theta)]^2$$

定理 2　对所有随机生成的训练集 θ 有:

$$PE^{**}\leqslant\bar{\rho}PE^* \quad (2)$$

式(2)中 $\bar{\rho}$ 是在 θ 与 θ'相互独立的情况下,残差 $Y-h(X,\theta)$和 $Y-h(X,\theta')$的加权相关系数。

上述定理给定了精确随机森林的前提:残差间的相关系数低以及错误决策树数目较少。为降低决策树的平均误差,随机森林回归选择对相关系数 $\bar{\rho}$ 加权处理。

随机森林回归算法的具体步骤可概括为

步骤 1:使用 bootstrap 方法对样本集进行重采样,进而随机生成 k 个训练集 $\theta_1,\theta_2,\cdots,\theta_k$。依据 k 个训练集进一步生成与之相对用的决策树$\{T(x,\theta_1)\}$,$\{T(x,\theta_2)\}$,…,$\{T(x,\theta_k)\}$。

步骤 2:从所有 M 个特征中随机生成 m 个特

征,并将其作为现下决策树分裂时的特征集。分裂方式则选择这 m 个特征中的最优分裂方式(通常来说,在随机森林构建过程中,m 的值不发生变化)。

步骤3:不对单棵决策树进行剪枝,令其最大程度生长。

步骤4:通过观测叶节点 $l(x,\theta)$ 的值并取平均可以获得面对新数据单棵决策树 $T(\theta)$ 的预测结果。

假定一个不为0且包含于叶节点 $l(x,\theta)$ 的观测值 X_i,权重 $w_i(x,\theta)$ 表示为

$$w_i(x,\theta)=\frac{1\{X_i\in R_l(x,\theta)\}}{\#\{j:X_j\in R_l(x,\theta)\}},(i=1,2,\cdots,n) \tag{3}$$

式(3)中的权重和为1。

步骤5:单棵决策树的预测值是通过因变量的观测值 $Y_i(i=1,2,\cdots,n)$ 加权平均得到的。单棵决策树的预测值表示为

$$u(x)=\sum_{i=1}^{n}w_i(x,\theta)Y_i$$

步骤6:对决策树的权重 $w_i(x,\theta_t)(t=1,2,\cdots,k)$ 取均值用以表示每个观测值 $Y_i\in(1,2,\cdots,n)$ 的权重 $w_i(x)$:

$$w_i(x)=\frac{1}{k}\sum_{i=1}^{k}w_i(x,\theta_t)Y$$

则随机森林回归算法的预测值表示为

$$u(x)=\sum_{i=1}^{n}w_i(x)Y_i$$

2 变量选取与数据说明

2.1 影响因素确定

(1)大气污染物浓度。大气污染物浓度是影响城市空气质量的直接因素,也是当前国际社会常用的城市空气质量评价指标,且各国关注的污染物种类和浓度取值时间差异较小[15]。依据国家《环境空气质量标准》(GB 3095-2012),确定了包括PM2.5、PM10、SO_2、NO_2、O_3、CO在内的6项污染物浓度作为城市空气质量影响因素。

(2)气象参数。人类生活生产会产生污染物进而排入大气中进而影响城市空气质量,然而当污染物的排放量相对平衡的状态下,城市空气质量依然会存在差异,这是由于气象参数的不同导致大气污染物进行沉降、传输、凝聚或者稀释。

选取了平均气温、最高气温、最低气温、平均相对湿度、平均风速、最大风速风向、日照时数、降水量、平均气压这9种气象参数作为城市空气质量影响因素,相关数据均来源于中国天气网历史气象数据。表1为2018-01-01—2020-07-31重庆市空气质量指数(AQI)与9种气象参数的相关系数,结果表明,城市空气质量与9种气象参数存在显著相关关系。

表1 空气质量指数与气象参数的相关关系

Table 1 Correlation between air quality index and meteorological parameters

气象参数	平均气温	最高气温	最低气温	平均气压	平均风速	最大风速风向	日照时数	降水量	平均相对湿度
相关系数	0.13	0.18	0.05	-0.07	-0.14	-0.11	0.46	-0.27	-0.47

(3)时间参数。同一城市在不同季节下的空气质量也会有所差异,冬夏两季相比于春秋季节需要更多地使用空调、暖气等,因此在预测城市空气质量时应当考虑季节因素。

2.2 数据预处理

在上述所确定的影响因素中,最大风速风向与季节均属于非数值型因素,对此,对其进行了量化,将非数值型因素转化为离散的数值型因素。之所以这样处理,是因为随机森林算法对数据的单位及量纲表现并不明显,也不需要对整理好的数据进行归一化处理,这也是选取该算法建立模型的原因之一。

(1)最大风速风向数据处理。将风向方位分为17类,分别为北、北偏东、东北、东偏北、东、东偏南、东南、南偏东、南、南偏西、西南、西偏南、西、西偏北、西北、北偏西及无风。并对其取值为[1,2,3,…,15,16,17]。

(2)季节数据处理。季节的取值为[1,2,3,4],分别代表春夏秋冬4个季节。

2.3 数据来源

最终选取影响因素共16项(表2),所使用的数据为2018-01-01—2020-07-31日重庆市空气质量监测站历史数据与历史气象数据。其中空气质量监

测站数据来自国家环保局,包含大气污染物六项因素与空气质量指数(AQI);历史气象数据来自中国天气网,包含 9 项气象参数。

表 2　影响因素选取

Table 2　Selection of influencing factors

特征类型	具体特征	取值范围	变　量
大气污染物	PM2.5	Real	0
	PM10	Real	1
	SO2	Real	2
	NO2	Real	3
	O3	Real	4
	CO	Real	5
气象参数	平均气温	Real	6
	最高气温	Real	7
	最低气温	Real	8
	平均相对湿度	Real	9
	平均风速	Real	10
	最大风速风向	[1,2,…,17]	11
	日照时数	Real	12
	降水量	Real	13
	平均气压	Real	14
时间参数	季节	[1,2,3,4]	15

3　预测模型的构建

基于随机森林回归的城市空气质量预测模型的整体思想是:确定影响城市空气质量的特征因素并收集整理数据集,然后应用随机森林回归进行模型构建,通过调整参数的最佳组合,不断优化模型。

3.1　测试集与训练集划分

共选取 2018-01-01—2020-07-31 共 943 条相关指标数据作为模型样本集。其中训练集与测试集样本比例为 8∶2,训练集样本 756 条,测试集数据样本 188 条。

3.2　网格搜索法参数寻优

随机森林算法性能的影响因素主要有两个:构建决策树时所用特征的数目及随机森林中决策树的数目,不同的参数选择会得到的预测结果与精准度也会不尽相同。对此,使用网格搜索法进行最优参数选取。网格搜索法的本质是指定参数值的穷举搜索方法,即将各参数的可能取值进行排列组合形成网格,进而使用交叉验证对网格中的所有点的表现进行评估,从而寻找出最优参数。具体步骤如下:

步骤 1:设定随机森林决策树棵数范围[1,160];决策树最大特征数范围[1,16]。

步骤 2:考虑到计算量,将决策树棵数的寻优参数步长设置为 10,决策树最大特征数的寻优参数步长设置为 1。

步骤 3:采用 Python 默认的 5 折交叉验证,其中 4 份作为训练数据,剩下的一份作为验证数据,从而生成不同的参数组合。

步骤 4:求不同参数组合在验证集上的测试误差,选取测试误差最小的参数组合作为最终参数。

通过对重庆市 2018-01-01—2020-07-31 的指标数据通过网格搜索法进行参数寻优后,共得到 240 组参数组合。部分参数组合结果见表 3。

表 3　部分参数组合及其测试误差

Table 3　Partial parameter combination and its test error

最大特征数	决策树数	测试误差
15	31	-4.91148
15	41	-5.09676
15	51	-4.98795
15	61	-5.23169
15	71	-4.50258
15	81	-5.14448
15	91	-4.69466
15	101	-5.02943
15	111	-4.61298

各参数组合中,测试误差最小的组合为决策树数目 71,决策树的最大特征数目 15。因此将其作为随机森林算法的最终参数。

3.3　预测结果度量指标

采用通用的 R^2(决定系数)、D_{MSE}(均方误差)、D_{MAE}(平均绝对误差)作为度量指标,进行基于随机森林回归的城市空气质量预测模型的性能分析。

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - f(x_i))^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{4}$$

$$D_{\mathrm{MSE}}=\frac{1}{n}\sum_{i=1}^{n}(y_i-f(x_i))^2\times 100\% \qquad (5)$$

$$D_{\mathrm{MAE}}=\frac{1}{n}\sum_{i=1}^{n}|y_i-f(x_i)|\times 100\% \qquad (6)$$

式(4)、(5)、(6)中,y_i 为真实值,$f(x_i)$ 为预测值,$\bar{y}$ 为真实值的均值,n 为样本数。其中,R^2 表示自变量解释的变异程度占总的变异程度的比例,它的值越接近 1 表示模型的准确度越高;D_{MSE}、D_{MAE}反映的是预测误差的大小,它们的值越小,越表明模型的预测精度。

4 预测及结果分析

在 Python 环境下,采用构造决策树为 71、特征数为 15 的最优参数组合对训练集进行训练,利用测试集对训练好的模型进行城市空气质量预测。各影响因子在模型中的重要程度见图 1。

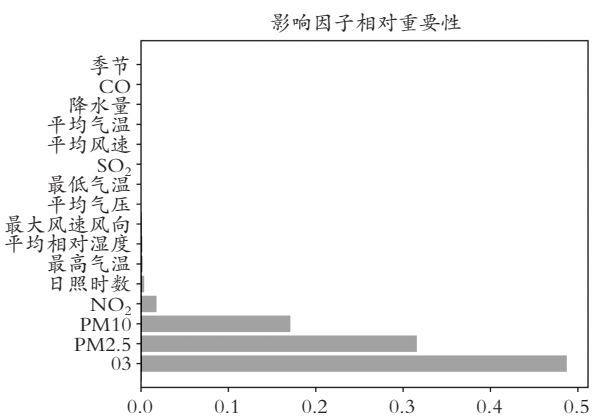

图 1 各影响因子相对重要性

Fig. 1 Relative importance of each influence factor

通过图 1 可以看出,O_3、$PM_{2.5}$、PM_{10}、NO_2、最高气温、日照时数、平均气温这几项因素的重要性程度较高,说明污染物浓度、气温、日照对城市空气质量的影响相对较大;相反,季节、CO、降水量、平均气温、平均风速等因素对城市空气质量的影响相对较小。

图 2 为模型预测值与实际值散点图,其中蓝色圆形点为空气质量预测值,黑色星形点为空气质量实际值。模型的预测值与实际值基本相吻合,但存在少数空气质量实际值偏高情况下的预测偏差。图 3 展示了模型预测结果和实际值的线性拟合效果。

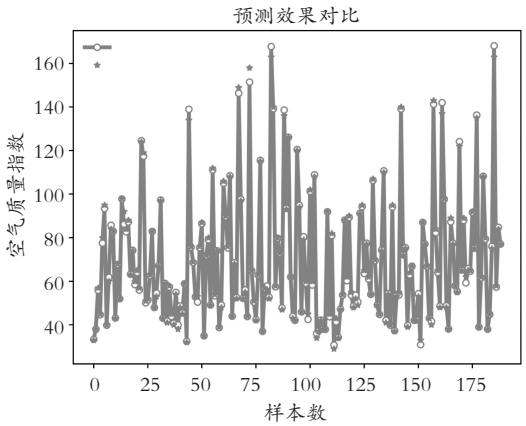

图 2 模型预测值与实际值散点

Fig. 2 Scatter of model predicted and actual values

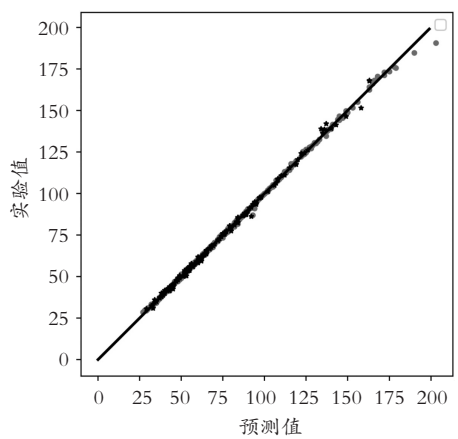

图 3 模型预测结果和实际值拟合效果

Fig. 3 Fitting effect of model prediction result and actual value

此外,分别计算训练集与测试集的 R^2(决定系数)、D_{MSE}(均方误差)、D_{MAE}(平均绝对误差),见表 4。

表 4 模型性能分析

Table 4 Model performance analysis

数据集	R^2	D_{MSE}	D_{MAE}
训练集	0.999	0.549	0.350
测试集	0.998	1.788	0.851

其中无论是训练集还是测试集,模型的确定性系数 R^2 都达到 99% 以上,说明模型具有较好的学习能力与泛化能力;就误差而言,D_{MSE} 与 D_{MAE} 在训练集与测试集上的取值均在可接受范围内。总体来说,提出的基于随机森林回归与气象参数的城市空气质量预测模型具有运行速度快、预测误差小、具有较高的预测精度等优点,具备较好的学习能力与泛化能力。

5　结束语

伴随当下人工智能技术的不断发展、大数据时来临,传统统计学方法劣势凸显,面对海量气象数据愈多学者选择使用机器学习方法构建城市空气质量预测模型。在此背景下,综合考虑污染物浓度、气象参数、时间参数等多方面影响因素,通过网格搜索法调整参数的最优组合,构建基于随机森林回归算法的城市空气质量预测模型,并以重庆市 2017-01-01—2020-07-31 的指标数据进行实证,结果表明,在模型下,训练集与测试集的确定性系数 R^2 都达到 99% 以上,证实了模型具有运行快速、预测准确、不易过度拟合等优点。此外,针对预测中出现的高值空气质量预估偏差问题,是下一步的研究内容。

参考文献(References):

[1] 沈劲,钟流举,何芳芳,等. 基于聚类与多元回归的空气质量预报模型开发[J]. 环境科学与技术,2015,38(2):63—66
SHEN J,ZHONG L J,HE F F,et al. Development of Air Quality Prediction Model Based on Clustering and Multiple Regression [J]. Environmental Science and Technology, 2015,38(2):63—66(in Chinese)

[2] 李博群,贾政权,刘利平. 基于模糊时间序列的空气质量指数预测[J]. 华北理工大学学报(自然科学版),2018,40(3):78—86
LI B Q,JIA Z Q,LIU L P. Prediction of Air Quality Index Based on Fuzzy Time Series [J]. Journal of North China University of Technology (Natural Science Edition), 2018,40(3):78—86(in Chinese)

[3] 汤泽梅,我国部分城市空气质量指数的聚类、建模及预测研究[D]. 昆明:云南师范大学,2018
TANG Z M. Clustering, Modeling and Prediction of Air Quality Index of Some Cities in China [D]. Kunming: Yunnan Normal University, 2018(in Chinese)

[4] 王建书,王瑛,赵敏娴,等. ARIMA 模型在苏州市空气质量指数预测中的应用[J]. 公共卫生与预防医学,2019,30(2):18—20
WANG J S,WANG Y,ZHAO M X,et al. ARIMA Model in Suzhou City Air Quality Index Prediction Application [J]. Public Health and Preventive Medicine,2019,30(2):18—20(in Chinese)

[5] 司志娟,孙宝盛,李小芳. 基于改进型灰色神经网络组合模型的空气质量预测[J]. 环境工程学报,2013,7(9):3543—3547
SI Z J,SUN B S,LI X F. Air Quality Prediction Based on Improved Grey Neural Network Combination Model [J]. Journal of Environmental Engineering, 2013, 7(9): 3543—3547(in Chinese)

[6] 赵李明. 基于遗传算法和 BP 神经网络的广州市空气质量预测与时空分布研究[D]. 南昌:江西理工大学,2016
ZHAO L M. Guangzhou Air Quality Prediction and Spatial-Temporal Distribution Based on Genetic Algorithm and BP Neural Network[D]. Nanchang: Jiangxi University of Technology, 2016(in Chinese)

[7] 张楠,王鹏,白艳萍,等. 基于 MGWO-SVR 的空气质量预测[J]. 数学的实践与认识,2018,48(8):159—165
ZHANG N, WANG P, BAI Y P, et al. Air Quality Prediction Based on MGWO-SVR [J]. Practice and Understanding of Mathematics, 2018,48(8):159—165 (in Chinese)

[8] 夏润,张晓龙,基于改进集成学习算法的在线空气质量预测[J]. 武汉科技大学学报 2019,42(1):61—67
XIA R,ZHANG X L. Online Air Quality Prediction Based on Improved Ensemble Learning Algorithm [J]. Journal of Wuhan University of Science and Technology 2019,42(1): 61-67(in Chinese)

[9] 郑洋洋,白艳萍,续婷. 基于 SARIMA-SVR 组合模型的空气质量指数预测[J]. 河北工业科技,2019,36(6):436—441
ZHENG Y Y,BAI Y P,XU T. Prediction of Air Quality Index Based on SARIMA-SVR Combined Model[J]. Hebei Industrial Science and Technology, 2019,36(6): 436—441(in Chinese)

[10] 徐旭冉,涂娟娟. 基于决策树算法的空气质量预测系统[J]. 电子设计工程,2019,27(9):39—42
XU X R,TU J J. Air Quality Prediction System Based on Decision Tree Algorithm[J]. Electronic Design Engineering, 2019,27(9):39—42(in Chinese)

[11] 孔丽英,林晓玲,吴铮涛,等. 基于企业进销项发票数据的税收风险预测模型[J]. 肇庆学院学报,2020,41(2):1—8
KONG L Y,LIN X L,WU Z T,et al. Tax Risk Prediction Model Based on Enterprise Sales Invoice Data[J]. Journal of Zhaoqing University,2020,41(2):1-8(in Chinese)

[12] SANJIBAN S R, ROHAN C, KUN C L. Gradient Boosted Machines and Deep Neural Network for Stock Price Forecasting: A Comparative Analysis on South Korean Companies[J]. International Journal of Ad Hoc and Ubiquitous Computing, 2020, 33(1): 1105—1115

[13] KOUTAROU M, YASUNOBU N, HIDEHISA S, et al. Stroke Prognostic Scores and Data-Driven Prediction of Clinical Outcomes After Acute Ischemic Stroke[J]. Stroke, 2020, 51(5)

[14] 马冉,刘洪斌,武伟. 三峡库区草堂河流域土壤pH空间分布预测制图[J]. 长江流域资源与环境,2019,28(3):691-699
MA R, LIU H B, WU W. Spatial Distribution Prediction Mapping of Soil PH in Caotang River Basin of Three Gorges Reservoir Area [J]. Resources and Environment of Yangtze River Basin, 2019, 28(3): 691—699 (in Chinese)

[15] 王帅,潘本锋,张建辉,等. 环境空气质量综合指数计算方法比选研究[J]. 中国环境监测,2014,30(6):46-52
WANG S, PAN B F, ZHANG J H, et al. Comparison and Selection of Calculation Methods for Comprehensive Index of Ambient Air Quality[J]. China Environmental Monitoring, 2014, 30(6): 46—52 (in Chinese)

Urban Air Quality Prediction Model Based on Random Forest Regression and Meteorological Parameters: Take Chongqing as an Example

XU Yan-ping, CHEN Yi-an

(School of Mathematics and Statistics, Chongqing Technology and Business University, Chongqing 400067, China)

Abstract: In order to effectively predict urban air quality, promote urban air pollution prevention and control, and make up for the deficiency of low accuracy and poor fault tolerance of traditional statistical models for urban air quality prediction under the background of big data era, a prediction model of urban air quality based on Stochastic Forest regression is proposed. Considering the pollutant concentration, meteorological parameters, time parameters and other factors, the optimal combination of parameters was adjusted by grid search method, and the urban air quality prediction model based on Stochastic Forest regression algorithm was established. Based on the index data of Chongqing from January 1, 2017 to July 31, 2020, the air quality in Chongqing is predicted and analyzed. The results show that the certainty coefficients of training set and test set are above 99%, and the mean square error and average absolute error under the model on the training set and test set are within the acceptable range, which proves that the model has the advantages of fast running speed, small prediction error, high prediction accuracy, and good learning ability and generalization ability.

Key words: random forest regression; air quality prediction; meteorological parameters; air quality index

责任编辑:田 静

引用本文/Cite this paper:

徐艳平,陈义安. 基于随机森林回归和气象参数的城市空气质量预测模型——以重庆市为例[J]. 重庆工商大学学报(自然科学版),2021,38(6):118—124

XU Y P, CHEN Y A. Urban Air Quality Prediction Model Based on Random Forest Regression and Meteorological Parameters: Take Chongqing as an Example[J]. Journal of Chongqing Technology and Business University (Natural Science Edition), 2021, 38(6): 118—124