文章编号:1674 - 6139(2024)04 - 0071 - 05

基于随机森林的大气污染物实时排放总量估计研究

翟秀英

(菏泽市生态环境局东明县分局,山东 菏泽 274500)

摘 要:为了提高大气污染物实时排放总量估计方法的估计效果,设计基于随机森林的大气污染物实时排放总量估计方法。为保证提高数据的质量,分别对数据进行数据清洗、降噪处理和标准化计算。根据不同数据的属性,提取数据的气象特征、时间特征和地形特征。利用随机森林算法,对提取的数据特征进行回归处理,从而生成大气污染物实时排放估计模型。对大气污染物排放量进行预测,并对其进行增量处理,计算相应的预测误差,实现对大气污染物实时排放量的估计。测试结果表明,和对比方法相比,设计方法估计误差平均值为 4.21 mg/m^3,估计效果较好。

关键词:随机森林;大气污染物;实时排放总量;估计方法;方法设计

中图分类号:X51 **文献标志码**:B

Real - time Estimation of Total Air Pollutants Emissions Based on Random Forests

Zhai Xiuying

(Dongming County Branch of Heze Ecological Environment Bureau, Heze 274500, China)

Abstract: In order to improve the real - time estimation of total air pollutants emission, this paper designs a real - time total emission estimation method of air pollutants based on random forest. In order to improve the quality of the data, data cleaning, noise reduction and standardization calculation are performed on the data respectively. According to the attributes of different data, the meteorological features, time features and terrain features of the data are extracted. The random forest algorithm is used to perform regression processing on the extracted data features, so as to generate a real - time emission estimation model of air pollutants. The emission of air pollutants is predicted, and the corresponding prediction error is calculated by incremental processing, so as to realize the estimation of real - time emission of air pollutants. The test results show that compared with the comparison method, the average estimation error of the designed method is 4.21 mg/m^3, and the estimation effect is better.

Key words: random forest; atmospheric pollutants; real time total emissions; estimation method; method design

前言

为了实时掌握城市的空气变化情况,针对恶劣天气制定相应的解决措施,保障人们的日常生活,需要对大气污染物实时排放总量进行估计[1]。在上述背景下,不少研究学者对估计方法展开了研究。文献[2]统计车辆在行驶过程中的油耗量,并在网联车轨迹重构的作用下,规划车辆行驶路径,并对在该行驶过程中车辆的排放量进行估计,但该方法应用成本较高。文献[3]采集一段时间内轻型车污染物的排放数据,并对采集的数据进行预处理,在 MOVES 模型本地化的作用下,构建相应轻型车排放模型,对轻型车的排放因子进行计算,但该方法估计精准度较差。文献[4]对船舶辅机大气污染物排放量展开估算研究,建立了基于燃油消耗的排放因子,基于燃油消耗量,估算广州港船舶停泊工况辅助发动机不同污染物气体的排放总量,该方法引擎功率

收稿日期:2024 - 01 - 02

作者简介:翟秀英(1973 -),女,大学本科,高级工程师,研究方向:环境保护工程。

较低的船舶大气污染物排放因子更高,但估算结果的准确性还有待提升。在以往研究的基础上,文章设计了基于随机森林的大气污染物实时排放总量估计。先采集一段时间内大气污染物的排放数据,并对采集到的数据进行预处理,再提取出相关的数据特征,利用随机森林算法,对其进行回归处理,从而构建相应的大气污染物排放模型,实现对大气污染物实时排放总量的估计。

1 大气污染物实时排放总量估计方法设计

1.1 大气污染物实时排放数据预处理及特征提取

为实现对大气污染物实时排放总量的估计,需要对获取的大气污染物实时排放数据进行预处理。在文章中,数据预处理的过程主要包括三个步骤,分别为数据清洗、降噪处理和标准化处理[5]。其中,数据清洗针对数据集中的无效数据,进行相应的数据处理。在获取数据的过程中,由于多种因素的影响,导致获取的数据中存在重复、缺失、冗余等情况,影响到数据的质量。因此,需要对数据进行清洗。在清洗的过程中,将上述无效数据进行删除或者弥补,保证数据的完整性,提高数据的质量[6]。降噪处理是去除原始数据中的噪声,为后续设计提供较为精确的数据支持。标准化处理是将所有数据保持在同一维度,减少因数据维度不同造成的数据计算压力,提高计算的速度[7]。在上述过程中,降噪处理和数据标准化处理的具体过程如式(1)所示:

$$\begin{cases} f(x) = \int f(u)\exp\left[-\dfrac{(u-x)^2}{2\sigma^2}\right]du \\ X' = \dfrac{X - X_{\min}}{X_{\max} - X_{\min}} \end{cases} \quad 式(1)$$

式(1)中,$f(x)$ 表示降噪处理的结果,$f(u)$ 表示降噪函数,σ 表示需要进行降噪处理数据的均值,u、x 表示降噪处理前后数据的变化参数,X' 表示数据标准化处理的结果,X 表示原始数据值,$X_{\max}$ 表示原始数据的最大值,$X_{\min}$ 表示原始数据的最小值。通过上述公式,完成对数据的降噪处理和标准化处理,提高了数据的质量,为后续的提取数据特征提供可靠的数据支持。在对数据特征进行提取时,考虑到影响大气污染物排放量的因素有很多,如果仅是从海量的数据中提取出相关特征,将会导致提取的数据特征具有一定片面性[8]。因此,在进行数据特征时,根据不同的数据属性,选择不同的特征提取方式。具体特征提取的过程见图 1。

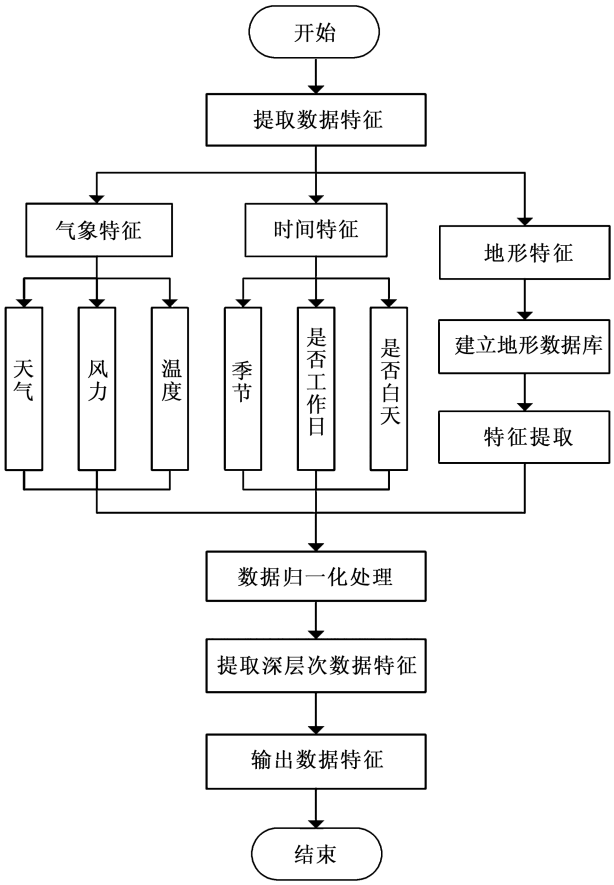

图 1 数据特征提取过程

如图 1 所示,考虑大气污染物的排放量会受到天气、所处区域和时间的影响,因此,在进行特征提取时,分别提取出数据的气象特征、时间特征和地形特征。其中,天气特征通过提取数据的温度、风力和天气来表示,时间特征则是当前数据所处的季节、是否为工作日、是否为白天来表示。在对上述两种数据特征进行提取时,考虑到提取的数据特征均为离

散型,因此,在特征提取后,需要对其进行归一化处理。对于地形特征来说,特征提取的过程较为复杂,需要先建立相应的地形数据库,并在其中存储大量的地形数据,以此为基础,分别提取出数据的相关特征,再对提取的数据特征进行归一化处理。为了获取更深层次的数据特征,将上述特征组合,为后续构建排放量估计模型奠定基础。

至此,完成对大气污染物实时排放数据预处理及特征提取的设计。

1.2 基于随机森林构建大气污染物排放量估计模型

以上述提取的数据特征作为基础,构建相应的大气污染排放量估计模型。文章利用随机森林算法构建模型,随机森林算法作为机器学习算法中的一种,能够将多个学习预测模型组合起来,对样本数据进行回归处理。和其他机器学习算法不同,随机森林算法的计算能力更强,能够处理海量高维度的数据,优化的参数也相对更少,能够有效提高大气污染物排放量估计模型的泛化性。

以提取的数据特征作为基础,构建相应的回归模型。在回归模型内部节点选择属性时,先将节点进行分裂处理,从分裂后的节点中提取出相应的分裂特征,为后续构建估计模型奠定基础。在上述过程中,随机森林算法具体应用如式(2)所示:

$$M = \frac{1}{n}\sum_{i=1}^{n}(f(x_i) - y_i)^2 \qquad 式(2)$$

式(2)中,M 表示构建的回归模型,n 表示提取的数据特征,$f(x_i)$ 表示节点分裂后的特征值,y_i 表示初始特征值。通过上述公式,得到相应的随机森林回归模型。在上述基础上,构建相应的大气污染物排放量估计模型。具体构建结果如式(3)所示:

$$\begin{cases} q = N(0,\varphi_n^2 + W(i)) \\ k(W_1,W_2) = \varphi_q^2 \exp\left(-\frac{1}{2\varphi^2}\left|W_1 - W_2\right|\right)^2 \cdot M \end{cases} \qquad 式(3)$$

式(3)中,q 表示数据的特征参数,$N(\ \ )$ 表示特征参数的估计核函数,φ_n 表示特征参数的分布函数,$W(i)$ 表示需要进行处理的数据,$k(W_1,W_2)$ 表示构建的大气污染物排放量估计模型,W_1、W_2 分别表示经过处理的特征参数值,φ_q 表示大气污染物排放量的规模参数,φ 表示大气污染物排放量的时间参数。通过上述公式,得到大气污染物的排放量估计模型。

至此,完成基于随机森林的大气污染物排放估计模型的构建。

1.3 实现对大气污染物实时排放总量的估计

在上述设计的基础上,实现对大气污染物实时排放总量的估计。具体估计流程见图 2。

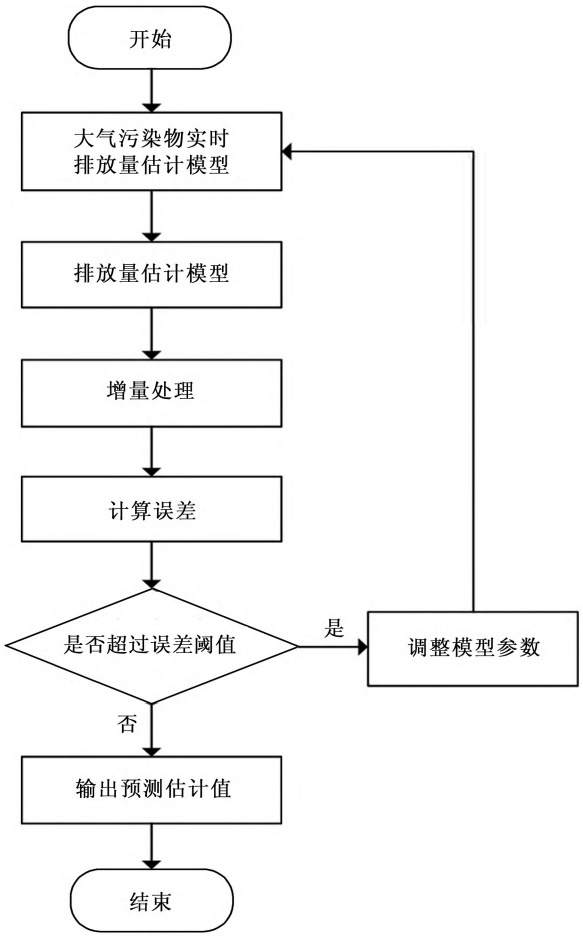

图 2 大气污染物实时排放总量估计流程

如图 2 所示,将上述构建的大气污染物实时排放量估计模型作为基础,对大气污染物的实时排放量进行预测,根据预测结果,先对其进行增量处理,

再计算其与实际排放量的误差。具体计算过程如式(4)所示：

$$h = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|k(x_i) - a_i\right|}{a_i} \quad \text{式(4)}$$

式(4)中，h 表示计算的预测误差，$k(x_i)$ 表示进行增量处理后的预测值，a_i 表示实际值。通过上述公式，计算出预测结果的误差，若该误差在设定的阈值范围内，则完成对大气污染物实时排放量的估计。若该误差过大，超过了设定的阈值，则需要调整模型参数，重新进行预测。

至此，完成基于随机森林的大气污染物实时排放总量估计方法的设计。

2 实验测试

2.1 实验准备

为验证文章设计的基于随机森林的大气污染物实时排放总量估计方法在实际应用中的性能，进行相关实验测试。

在此次实验中，从多个大气污染物排放相关网站上采集相关数据，作为此次实验数据，再对采集到的数据进行预处理。再对其进行特征提取，并将提取到的特征，利用随机森林算法进行回归处理。此外，在此次实验中，为保证实验结果的可靠性，设置了相应的对照实验。其中，文章设计的基于随机森林的大气污染物实时排放总量估计方法为方法 1，基于粒子群优化算法的大气污染物实时排放总量估计方法为方法 2，基于 BP 神经网络额大气污染物实时排放总量估计方法为方法 3。

2.2 实验结果与讨论

为对比三种方法在实际应用中效果，此次实验以三种方法估计误差作为评价指标，对比三种估计方法的性能。实验中，利用三种方法估计 20 天内大气污染物的排放总量，并与实际排放总量相比，统计排放误差。具体统计结果见表 1。

表 1 三种方法的估计结果

时间/天	实际排放量/(mg/m³)	三种方法的估计结果/(mg/m³)		
		方法 1	方法 2	方法 3
1	218.23	211.23	228.23	258.23
2	225.32	221.32	235.32	255.62
3	236.24	232.41	226.24	276.14
4	242.23	239.38	249.23	262.13
5	232.14	230.08	242.54	212.64
6	222.36	225.36	232.66	292.16
7	212.69	214.36	242.89	282.68
8	228.36	227.18	218.39	258.46
9	219.67	220.32	229.65	259.87
10	224.58	221.58	214.68	264.38
11	228.23	227.34	238.23	258.88
12	214.32	215.25	235.35	255.69
13	236.48	232.36	266.64	286.14
14	242.78	239.22	259.13	272.13
15	212.14	210.08	242.63	282.64
16	208.42	205.36	232.42	292.23
17	224.69	223.36	242.11	282.24
18	223.37	227.18	208.39	298.26
19	242.69	240.32	219.65	289.67
20	238.56	241.58	264.69	274.39

如表 1 所示，利用三种方法对大气污染物排放总量进行估计时，方法 1 的估计值与实际值相差不多，方法 2 和方法 3 与实际值相差较大。其中，方法 1 与实际测量值的最大误差为 7.15 mg/m³，方法 2 的最大误差为 30.16 mg/m³，方法 3 的最大误差为 83.81 mg/m³。同时，从整体来看，方法 1 的平均误差为 4.21 mg/m³，方法 2 的平均误差为 22.36 mg/m³，方法 1 的平均误差为 65.48 mg/m³，由此可见，方法 1 的估计效果最好。因此，文章设计的基于随机森林的大气污染物实时排放总量估计方法在实际应用中效果最好，估计误差较小，能够较为准确地反映当前大气污染物的排放情况，为相关部门提供较为准确的数据支持。

3 结束语

为了改善以往的大气污染物实时排放总量估计方法存在的估计效果较差的问题,文章针对大气污染物实时排放总量估计方法的问题,设计了基于随机森林的大气污染物实时排放总量估计方法。通过对数据的清洗、降噪和标准化处理,提取了数据的气象、时间和地形特征,利用随机森林算法对特征进行回归处理,生成了相应的回归模型。利用该模型对大气污染物排放量进行预测,并计算增量和预测误差,实现了对大气污染物实时排放量的估计。实验测试表明,该方法相较于传统方法具有更好的估计效果,平均估计误差为 4.21 mg/m^3。因此,文章所设计的基于随机森林的大气污染物实时排放总量估计方法具有较高的实用价值,为大气污染物的监测和控制提供了有效手段。

参考文献:

[1]刘丹,李林山,曹平,等. 恩施州大气污染物排放量估算及清单构建[J]. 环境科学与技术,2021,44(1):207 -215.

[2]蒋阳升,刘梦,王思琛,等. 基于网联车轨迹重构的交通油耗和排放估计方法[J]. 安全与环境学报,2022,22(4):2147 -2155.

[3]单肖年,刘皓冰,张小丽,等. 基于 MOVES 模型本地化的轻型车排放因子估计方法[J]. 同济大学学报(自然科学版),2021,49(8):1135 -1143;1201.

[4]何俊杰,陈鸿展,陈俊文,等. 基于实测的船舶辅机大气污染物排放量估算[J]. 环境科学学报,2021,41(12):5055 -5062.

[5]李辉,孙雪丽,庞博,等. 基于碳减排目标与排放标准约束情景的火电大气污染物减排潜力[J]. 环境科学,2021,42(12):5563 -5573.

[6]李佳硕,孙千惠,王文鑫,等. “上大压小”降低中国煤电的大气污染排放——基于供应链视角[J]. 中国环境科学,2023,43(4):2047 -2056.

[7]郭凤艳,杨飞,邓双,等. 山西省某市焦化行业大气污染物排放特征[J]. 环境科学研究,2021,34(12):2887 -2895.

[8]丁镭,刁贝娣. 空间计量经济学视角下的浙江工业大气污染物排放及社会影响因素[J]. 环境污染与防治,2021,43(1):132 -138.

美韩研究团队开发出一种耐疲劳电解质膜

韩国仁川国立大学与哈佛大学联合研究团队成功开发出一种耐疲劳的电解质膜。

研究团队创造了一种由 Nafion 和全氟聚醚(PFPE)组成的互穿网络电解质膜。Nafion 是一种常用的具有质子导电性的塑料电解质,PFPE 则形成了一种耐用的橡胶聚合物网络,这种橡胶的加入虽然略微降低了电化学性能,但显著提高了耐疲劳阈值和使用寿命。研究团队测试了不同 PFPE 比例的电解质膜,发现 50% 饱和度的 PFPE 电解质膜表现出良好的电化学性能,与原始的 Nafion 膜相比,这种 Nafion - PFPE 膜将疲劳阈值提高了 175%,并将燃料电池的使用寿命延长了 1.7 倍。此外,未改性的 Nafion 膜的使用寿命为 242 小时,而复合膜的使用寿命达到了 410 小时。该研究在众多领域具有重要意义,相关研究成果发表于《Advanced Materials》。

本文摘自国外相关研究报道,文章内容不代表本网站观点和立场,仅供参考。

来源:中华人民共和国科学技术部网站

https://www.most.gov.cn/gnwkjdt/202403/t20240314_189965.html