

基于三种机器学习模型的太湖总氮浓度预测

赵 朔¹, 桂峰兰², 柳后起³

(1. 辽宁省生态环境保护科技中心, 沈阳 110161; 2. 中共辽宁省委党校, 沈阳 110004;
3. 中国科学技术大学苏州高等研究院, 江苏 苏州 215123)

摘要: 市政污水处理厂出水是自然水体的重要污染源, 处理后的生活污水排入水体, 对受纳水体水质产生重要的影响。为实现对太湖总氮浓度的准确预测, 收集并分析了太湖水体水质监测数据和太湖流域212个污水处理厂的实时运行监测数据, 采用皮尔逊相关系数(Pearson correlation coefficient)分析了太湖水体总氮浓度与市政污水处理厂相关运行指标的相关性, 结合相关性较高的前五项目标与太湖水质监测数据, 利用临近算法(KNN), 决策树以及AdaBoost三种机器学习模型对太湖水体总氮浓度月平均值进行了预测。其中AdaBoost的精度更高、准确性更好, 拟合优度为0.84, 平均绝对误差在14.08%以内。模型特征重要性分析表明, 太湖硝态氮、氨氮和总磷等指标对总氮浓度预测有重要的影响。

关键词: 皮尔逊相关系数; 市政污水处理厂; N排放; AdaBoost机器学习

中图分类号: TV11 **文献标识码:** A

赵 朔, 桂峰兰, 柳后起. 基于三种机器学习模型的太湖总氮浓度预测[J]. 中国农村水利水电, 2022(6): 24-28, 39.

ZHAO S, GUI F L, LIU H Q. Prediction of nitrogen concentration in Taihu Lake based on AdaBoost machine learning model [J]. China Rural Water and Hydropower, 2022(6): 24-28, 39.

Prediction of Nitrogen Concentration in Taihu Lake Based on AdaBoost Machine Learning Model

ZHAO Shuo¹, GUI Feng-lan², LIU Hou-qi³

(1. Liaoning Science and Technology Center for Ecological and Environmental Protection, Shenyang 110161, China;

2. Party School of Liaoning Provincial Committee of the CPC, Shenyang 110004, China;

3. Suzhou Institute for Advanced Study, USTC, Suzhou 215123, Jiangsu Province, China)

Abstract: Domestic waste water is an important source of natural water pollution. The treated domestic waste water is discharged into natural water bodies through the pipe network, which has a serious impact on the quality of the receiving water body. The water quality monitoring data from Taihu Lake and the plant-resolved and actual operation data of 212 waste water treatment plants in Taihu Lake from 2007 to 2015 are collected and analyzed in order to accurately predict the total nitrogen in Taihu Lake. The Pearson correlation coefficient is used to calculate the relationship between Taihu Lake water quality and the WWTP effluent indexes. For the top five items with high correlation, three machine learning models, K-Nearest Neighbors (KNN), Decision Tree, and AdaBoost are used to predict the monthly average TN in Taihu Lake. In general, AdaBoost has higher precision and better accuracy, with a goodness-of-fit index of 0.84 and a root mean square error less than 14.08%, which establishes a good mathematical model for predicting of TN concentration of Taihu Lake. Meanwhile, the model finds that $\text{NO}_3\text{-N}$, $\text{NH}_4\text{-N}$, TP in Taihu Lake and $\text{NH}_4\text{-N}$ in effluent of WWTP can cause an important impact on the TN concentration of Taihu Lake.

Key words: Pearson correlation coefficient; domestic wastewater treatment plant; nitrogen emission; AdaBoost machine learning model

收稿日期: 2021-09-07

基金项目: 中国博士后科学基金资助项目(2018T110566)。

作者简介: 赵 朔(1981-), 男, 高级工程师, 主要从事环境科学研究。E-mail: 18253678@qq.com。

通讯作者: 柳后起(1981-), 男, 正高级工程师, 博士, 硕士生导师, 主要从事生态保护与水污染控制研究。E-mail: liuhq@ustc.edu.cn。

0 引 言

市政污水处理厂出水是河流湖泊重要的外来污染排放源之一,对受纳水体的水质存在重要影响。当前,人类活动彻底改变了正常的氮循环,众多流域氮负荷显著增加,很多地区氮含量超过了地表水限值标准(1.0 mg/L)。非 N_2 固定剂的大量使用,改变了湖泊营养收支和循环特征^[1]。此外,在我国大规模和迅速改善废水处理是普遍采用措施(如市政污水出水标准的不断提高),但这种对受纳水体营养状况的影响很少在广泛的空间与时间尺度上得到验证。研究表明,我国东部的主要湖泊由于总氮和总磷浓度的不同变化显示出氮磷质量比增加,实际上湖泊氮磷质量比率的增加与城市污水处理的快速改善有关^[2]。作为水体富营养化主要营养盐,氮已成为影响太湖东部大多数站点水质的主要因素^[3],研究太湖氮浓度对于太湖水质监测与评价水体富营养化尤为重要。水质预测主要有机理性和非机理性两种,由于水环境的多因子、多介质的特性,常规的机理性水质模型如 Streeter-Phelps (S-P) 一维稳态氧平衡模型^[4], WASP 富营养化模型^[5], SWAT 水质预测模型等^[6],这些方法多需要复杂的边界条件和大量完善的监测资料而难以推广应用,而非机理模型则是利用数学的方法进行模拟,将影响因子作为输入变量,被预测的指标作为输出变量进行分析同时又可以特征分析以评估影响因子的相关重要程度。如胡志洋^[7]等利用 BP 神经网络模型对太湖梅梁湾区域的叶绿素 a 浓度进行预测,成浩科^[8]等利用随机森林的方法对淮河流域河流总磷浓度进行预测,付泰然等^[9]利用自编码式 BP 神经网络对养殖水体的亚硝态氮进行预测,李修竹等^[10]人利用支持向量机的方法对长江口岸的叶绿素 a 浓度进行预测。这种水质预测模型具有自适应性强、容错率高等优点,此外也有着较高的预测精度^[7],在水质预测应用的同时还能有效确定各项影响指标的重要性关系。因此,本文基于 2007-2015 年期间太湖水体八个监测站点的月度水质监测数据,首先借助皮尔逊相关系数分析市政污水处理厂各项运行指标与太湖水体总氮浓度的相关性以确定 5 项主要的影响因素,再结合 16 种太湖水质监测指标输入到三种机器学习模型(决策树, KNN 和 AdaBoost)中对太湖流域总氮浓度进行预测。此外,从预测效果较好的模型中输出了 5 项重要性特征以揭示太湖总氮浓度的重要影响因素。

1 数据与方法

1.1 数据来源

公开的数据集来自太湖水域的 8 个常规监测站(图 1),数据集的时间跨度为 2007 年至 2015 年^[9],记录频率为每月一次。数据集主要包括透明度、pH、水深、水温、水位、电导率等指标,以及总磷(TP)、总氮(TN)、硝态氮(NO_3-N)、亚硝酸盐氮(NO_2-N)、氨氮(NH_3-N)、溶解氧(DO)等 10 个浓度指标。

此外,收集了江苏和浙江等 212 个太湖流域市政污水处理厂运行数据,时间跨度为 2007 年至 2015 年,频率为每月一次,主要指标包括污水处理量、进水量、出水量、污泥产量等 10 个质量信息指标以及进出水 COD、氨氮、总氮、总磷、悬浮物等 12 个浓度信息指标,共计 22 项指标。

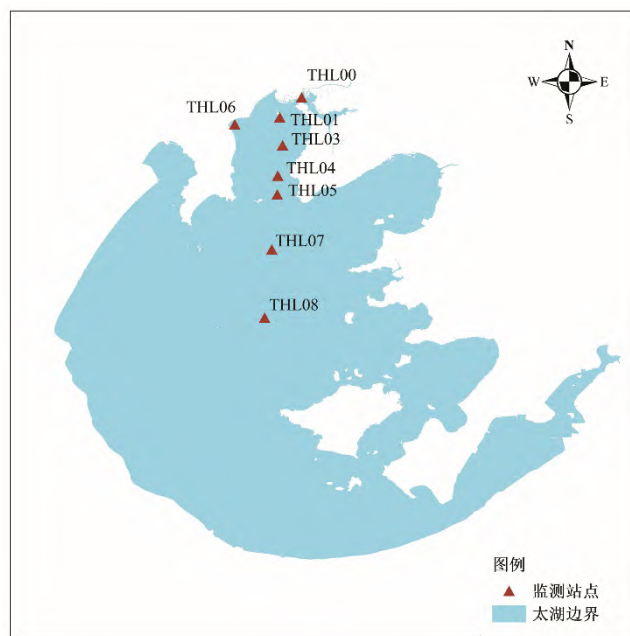


图 1 太湖水质监测站点分布^[11]

Fig.1 Distribution of water quality monitoring stations in Taihu Lake

1.2 基于皮尔逊相关系数法的指标筛选

皮尔逊相关系数本质为一种线性相关系数,相关系数的绝对值越大表明相关性越强。相关性系数越接近 1 或者 -1,相关度越强。越趋向于 0,相关度越弱,我们首先评估了市政污水处理厂 22 项运行指标与太湖流域总氮的相关性。

步骤 1 数据初始无量纲化。由于不同的监测数据量纲不同,在皮尔逊系数分析前对水质监测数据进行无量纲化处理,具体计算公式为:

$$x_i(k) = \frac{X_i(k) - \min X_i(k)}{\max X_i(k) - \min X_i(k)} \quad (1)$$

式中: i 为不同的监测指标类别; k 为不同类别对应的数据点。

步骤 2 计算离均差平方与离均差积和。

变量 X 的离均差平方和为:

$$l_{xx} = \sum (x_i - \bar{x}_i)^2 \quad (2)$$

变量 Y 的离均差平方和为:

$$l_{yy} = \sum (y_i - \bar{y}_i)^2 \quad (3)$$

X 与 Y 的离均差积和为:

$$l_{xy} = \sum (x_i - \bar{x}_i)(y_i - \bar{y}_i) \quad (4)$$

步骤 3 计算两个变量之间的皮尔逊相关系数,计算公式为:

$$r = \frac{l_{xy}}{\sqrt{l_{xx}l_{yy}}} \quad (5)$$

按照上述步计算两个变量之间的皮尔逊相关系数并对相关系数进行排序。

1.3 决策树原理

决策树(decision tree)是一类常见的、较为简单的机器学习方法,其目的是根据损失函数最小化的原则产生一棵泛化能力强,处理未见示例能力强的决策树^[12]。决策树方法具有易于理解和实现、分类规则简单符合人类思维方式等特点,研究人员已将其广泛应用于地震预警^[13]、震后损失评估、地质灾害评

价^[14]等诸多防震减灾领域。决策树在处理大样本数据时存在易过拟合、对连续性字段较难预测等问题。现有的决策树学习主要包含ID3算法,C4.5算法以及CART算法。

1.4 KNN模型原理

k近邻法(k-nearest neighbor, k-NN)是1967年由Cover T和Hart P提出的一种基本分类与回归方法^[15]。基本概念为:存在一个样本数据集合,所有特征属性已知,并且样本集中每个对象都已知所属分类。对未知分类的待测对象,将待测对象的每个特征属性与样本集中数据对应的特征属性进行比较,然后计算提取样本最相似对象(最近邻)的分类标签。一般只选择样本数据集中前k个最相似的对象数据,最后根据k个数据的特征和属性判断待测数据的分类。

1.5 AdaBoost模型原理

AdaBoost是一种迭代算法,核心思想是针对同一个训练集训练不同的分类器,即弱分类器,然后把这些弱分类器进行集合,构造一个更强的最终分类器^[16]。算法本身是通过改变数据分布来实现,它根据每次训练集中的每个样本的分类是否正确,以及上次的总体分类的准确率来确定每个样本的权值。将修改权值后的新数据传至下层分类器进行训练,然后将每次训练得到的分类器融合起来作为最后的决策分类器。AdaBoost是一种具有高精度、操作简便的分类器,不用考虑特征筛选和过拟合等问题。

步骤1输入训练集 $D = \{(x_1, y_1), (x_2, y_2), (x_3, y_3), \dots, (x_m, y_m)\}$, 其中 $x_i \in X$: 训练样本集合, $y_i \in Y = \{-1, 1\}$; 采用弱学习算法 WeakLearn; 迭代训练轮数 N 。初始化过程中令每个样本的初始权值为:

$$d_i(i) = \frac{1}{m} \quad (6)$$

步骤2迭代开始,并赋予当前分布 D_n 调用 WeakLearn, 得到弱分离器 h_n :

$$\text{for } n = 1 \text{ to } N \quad (7)$$

计算弱分离器 h_n 在当前分布下的错误率:

$$\varepsilon_n = \sum_{i: h_n(x_i) \neq y_i} d_n(i) \quad (8)$$

在迭代计算过程中,如果 $\varepsilon_n > 0.5$, 则令 $N=n-1$, 同时终止迭代。

步骤3结合结果计算分类器 h_n 在最终分类器集合中的加权系数:

$$\alpha_n = \frac{1}{2} \ln\left(\frac{1 - \varepsilon_n}{\varepsilon_n}\right) \quad (9)$$

同时更新样本分布:

$$D_{n+1}(i) = \frac{D_n(i)}{z_n} \times \begin{cases} \exp(-\alpha_n), h_n(x_i) = y_i \\ \exp(\alpha_n), h_n(x_i) \neq y_i \end{cases} = \frac{D_n(i) \exp[-\alpha_n y_i h_n(x_i)]}{z_n} \quad (10)$$

步骤4计算结果合并到模型输出函数结果:

$$H(x) = \text{sign}\left(\sum_{n=1}^N \alpha_n h_n(x)\right) \quad (11)$$

1.6 模型评价指标

本次评价指标主要包括拟合优度(R^2)和均方误差(RMSE)以及平均百分比误差(MAPE)。本次模型主要基于Python中的sklearn库来实现。

拟合优度 R^2 是指回归方程对观测值的拟合程度。度量拟合优度的统计量是判定系数 R^2 。 R^2 的取值范围是 $[0, 1]$ 。 R^2 的值越接近1, 说明回归方程对观测值的拟合程度越好; 反之则表明回归方程对观测值的拟合程度越差, 即:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - f_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} y_i \quad (12)$$

RMSE是均方根误差,指预测值与真实值偏差的平方和与观察次数比值的平方根,用来衡量预测值同观察值之间的偏差程度,即:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - f_i)^2} \quad (13)$$

MAPE是平均百分比误差,是误差与原始数据值相比较的过程,结果越趋向于0则为完美模型,大于100%则为劣质模型。即:

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - f_i|}{y_i} \quad (14)$$

式中: y_i 为真实值; f_i 为预测值。

2 结果与分析

2.1 皮尔逊系数相关性结果分析

利用市政污水处理厂的22项运行指标与太湖流域总氮浓度进行皮尔逊相关性分析,选取其中相关性较大的5项指标。皮尔逊系数的相关性绝对值直接体现了影响因素与结果之间的关联程度,其绝对值越接近于1表明指标间相关性越强。根据皮尔逊系数计算步骤分析了太湖水体总氮浓度与市政污水处理厂各出水排放指标的相关性,结果见表1。

由表1可知,经相关性分析后,市政污水处理厂出水氨氮浓度,污水排放总量,以及进出水总氮浓度与太湖总氮浓度有显著的影响,同时进水的BOD,COD,SS,以及总磷浓度有着较弱的相关性。因此,选取了污水处理厂的进出水的总氮浓度,氨氮浓度以及污水处理量等五项(表1)作为输入变量的一部分。

市政污水处理厂运行指标与太湖水体总氮相关性较高的主要原因是市政污水处理厂的尾水排放是太湖N输入的重要

表1 皮尔逊系数相关性结果

Tab.1 The result of Pearson coefficient analysis

序号	因子	相关性	序号	因子	相关性
1	出水 NH ₄ -N	0.67	6	进水 BOD	0.37
2	污水排放量	0.56	7	进水 BOD	0.33
3	进水 TN	0.51	8	进水 SS	0.32
4	出水 TN	0.50	9	进水 TP	0.32
5	进水 NH ₄ -N	0.45	10	出水 COD	0.27

注: NH₄-N 为氨氮; TN 为总氮; BOD 为生化需氧量; COD 为化学需氧量; SS 为固体悬浮物; TP 为总磷。

来源之一,实际上,环太湖流域的市政污水处理厂最早开始严格执行《城镇污水处理厂污染物排放标准》中的一级A标准^[17]。尽管如此,一级A标准所规定的主要污染物的排放标准仍远远高于地表水V类水要求的各项指标,例如一级A标准中规定的总氮浓度是地表水V类水体规定中的7.5倍,氨氮浓度是其2.5倍。因此,尽管市政污水处理厂的出水排放严格达到了1级A标准,却仍然是湖,库水体的富营养物质的重要污染源^[18]。此外,污水排放的总氮和氨氮是影响太湖水体总氮浓度的直接因素,这也是市政污水中氨氮和总氮的相关指标高于其他指标(例如COD、BOD)的主要原因。

2.2 三种机器学习模型构建

2.2.1 模型训练

依据上述皮尔逊结果,模型输入变量为从污水处理厂选取的五项指标以及太湖的水质监测数据集,输出为太湖水体的8个站点的月度均值总氮浓度。采用将随机选取80%的数据量为训练数据集,20%作为测试数据集,同时利用临近算法(KNN)和决策树以及AdaBoost三种模型进行训练。实验采用Python编程语言实现算法。

2.2.2 模型训练与测试结果

利用已经训练好的模型对测试集数据值进行测试,三种机器模型的实测值与预测值对比如图2,模型的性能指标结果见表2。

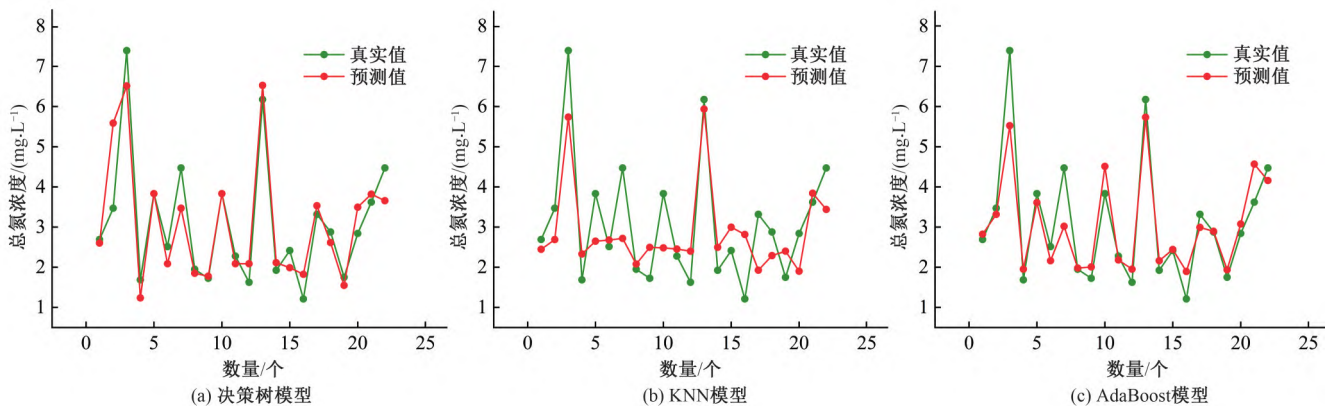


图2 模型决策树(Decision Tree Regressor), KNN和AdaBoost的训练输出与监测值的对比

Fig.2 Comparison between and measured values of TN and predicted values of DTR, KNN and AdaBoost Model.

从同类型的总氮浓度预测模型横向对比来看(表2),BP神经网络有着最高的拟合优度(R^2),这主要是由于神经网络具有高度的非线性,可以更好拟合非线性关系,以及反向传播机制可以不断将误差反传回来以调整网络来修改各层神经元的权值直至达到期望目标^[22]。而AdaBoost有着最低的平均绝对误差和均方根误差,在回归预测类问题上具有更小的误差,体现出更高的预测准确度,明显优于其他模型。相比于最好的BP神经网络, $RMSE$ 数值也由0.838下降至0.832。

2.2.3 模型重要性特征提取

对于输入的综合指标(5个相关性较高的市政污水处理厂相关指标以及16项太湖水质监测指标),分别从结果较好的决策树和AdaBoost模型中提取了前5个相关性特征指标(图3),相关性特征对模型预测结果有重要影响的特征,因此输出并分析了前5个重要特征。

表2 模型输出结果

Tab.2 performances of all Methods

模型	R^2	RMSE	MAPE/%
KNN	0.606	0.933	29.680
决策树	0.627	0.908	18.091
Prophet ^[21]	-	-	20.770
多元线性回归 ^[22]	0.866	1.059	-
BP神经网络 ^[22]	0.931	0.838	-
AdaBoost	0.836	0.602	14.081

由图2可知,相较于KNN和决策树,AdaBoost模型更好的反映了太湖总氮浓度的变化趋势,其平均百分比误差为14.081%,而KNN整体预测结果较为平缓,与实际浓度变化情况相差较大,有着最大的平均百分比误差29.680%,模型准确度远远低于其他两种模型。此外,AdaBoost模型拟合优度(R^2)达到0.836,表明模型更好地拟合了太湖总氮浓度变化情况。整体来看,相较于传统的机器学习模型KNN和决策树,AdaBoost模型可以更准确地预测市政污水处理厂出水排放对太湖流域氮浓度的影响。在特征种类和特征数目一定的情况下,由多种算法组成的集成学习有着更加明显的优势,不论是尺度的地表水质预测^[19],或者是河流水质预测^[20],长江断面水质预测等^[21],集成学习模型均具有更好的预测准确度和较低的拟合误差,模型预测效果得到了极大的提升。

结果表明太湖自身氮负荷(太湖水体硝态氮浓度和氨氮浓度)是影响太湖总氮浓度的重要特征,这在决策树和AdaBoost模型中均得到了体现。尤其在决策树模型中,硝态氮浓度这一指标的重要性远远大于其他特征,在AdaBoost模型中两者也明显高于其他特征。说明太湖水体自身氮负荷对总氮浓度变化有着更显著的影响。在一些空间回归预测模型中也有类似的研究结论,水体自身的氮负荷是影响总氮浓度的显著因素^[23],然而鉴于我们模型的输入数据的局限性,模型无法注意到土地类型,人口密度和土壤径流等因素^[23]。值得关注的是,太湖中磷负荷对总氮也有一定的影响。这可能是由于水体的生物对氮磷富营养物质的同时利用所致^[24]。

此外,太湖同时接受来自市政污水处理厂的点源污染,而这些重要的点源污染也是重要的N负荷来源之一^[23]。机器学习模型也考虑到了来自于市政污水的处理体积(AdaBoost,第三

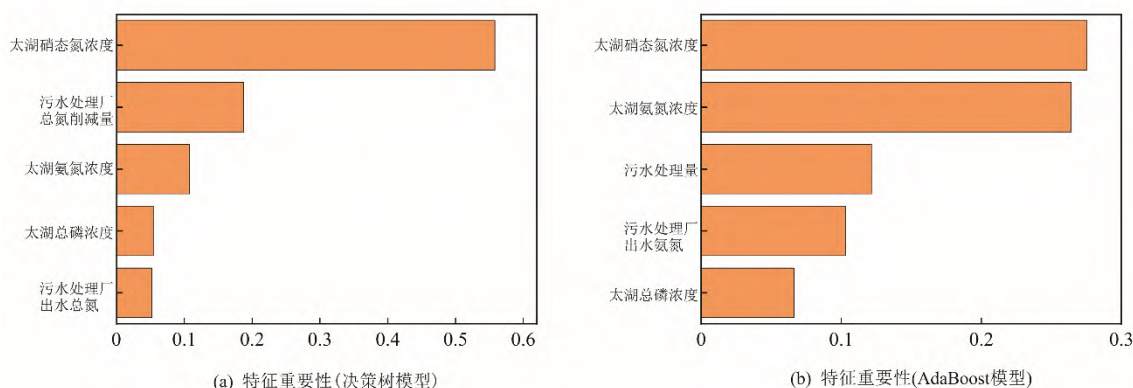


图3 模型的前5个重要特征

Fig.3 The importance of the top 5 important features

位)和出水氨氮的浓度(AdaBoost,第四位)是重要的影响特征指标。这说明市政污水处理厂出水氨氮是太湖流域重要的氮负荷来源^[18],同时市政污水处理厂的排放量也是重要的影响因素。由于在当前出水水质不断提高达到1级A后,市政污水处理厂尾水浓度也依然远高于地表水V类标准。同时日益增加的尾水排放量,使得排入太湖水体的N质量并没有明显减少,其绝对质量有增加趋势^[25]。整体来看,对于太湖流域总氮浓度影响最大的是太湖自身水体的氮负荷,此外是外N源的输入,如市政污水处理厂尾水中的氨氮等。

3 结 论

(1) 本研究利用KNN、决策树和AdaBoost三种机器学习模型分析了太湖水质的相关因素。三种模型中决策树模型的均方根误差和平均百分比误差分别比KNN降低2.68%和39.05%。而AdaBoost可以将RMSE和MAPE更进一步降低且呈现更好的拟合优度(约0.80),为高精度的水质预测提供借鉴。

(2) 模型特征重要性分析表明,对太湖流域总氮浓度影响最大的是太湖自身水体的硝态氮以及氨氮浓度,其次是污水排放的相关指标,如处理水量,出水氨氮浓度等。因此,为有效控制太湖氮浓度,既要控制污水处理厂出水氮浓度相关指标(氨氮,总氮等),也要对市政污水排放量加以控制。

(3) 为了控制或降低太湖流域总氮浓度,首先需要控制太湖水体自身的总氮浓度(硝态氮以及氨氮浓度),并需要减少氨氮的外源排入。

参考文献:

[1] YU C, HUANG X, CHEN H, et al. Managing nitrogen to restore water quality in China [J]. Nature, 2019, 567(7 749): 516-520.

[2] TONG Y, WANG M, PENUELAS J, et al. Improvement in municipal wastewater treatment alters lake nitrogen to phosphorus ratios in populated regions [J]. Proc Natl Acad Sci U S A, 2020, 117(21): 11 566-11 572.

[3] WANG J L, FU Z S, QIAO H X, et al. Assessment of eutrophication and water quality in the estuarine area of Lake Wuli, Lake Taihu, China [J]. Science of the Total Environment, 2019, 650: 1 392-1 402.

[4] FAN C, WANG W-S, LIU K F-R, et al. Sensitivity Analysis and Water Quality Modeling of a Tidal River Using a Modified Streeter-Phelps Equation with HEC-RAS-Calculated Hydraulic Characteristics [J]. Environmental Modeling & Assessment, 2012, 17(6): 639-651.

[5] 李 军,张明华. WASP富营养化水质模型在GIS中的完全集成研究[J]. 长江流域资源与环境, 2012, 21: 93-98.

[6] ZHANG Y, XIA J, SHAO Q, et al. Water quantity and quality simulation by improved SWAT in highly regulated Huai River Basin of China [J]. Stochastic Environmental Research and Risk Assessment, 2011, 27(1): 11-27.

[7] 胡志洋,李翠梅,薛天一. 基于灰色关联和ABC-BP神经网络的叶绿素a浓度预测[J]. 水电能源科学, 2021: 55-58.

[8] 成浩科,沈 菲. 基于随机森林的河流总磷预测模型及影响因素分析[J]. 环境保护科学, 2021: 62-67.

[9] 付泰然,刘广鑫,万全元,等. 基于栈式自编码BP神经网络预测水体亚硝态氮浓度模型[J]. 水产学报, 2019: 958-967.

[10] 李修竹,苏荣国,张传松,等. 基于支持向量机的长江口及其邻近海域叶绿素a浓度预测模型[J]. 中国海洋大学学报(自然科学版), 2019, 49(1): 69-76.

[11] 闵 岫,钱荣树,朱广伟,等. 2007-2015年太湖水体理化监测数据集[J]. 中国科学数据(中英文网络版), 2020, 5(1): 85-93.

[12] ZENKO B, TODOROVSKI L, DZEROSKI S. A comparison of stacking with meta decision trees to bagging, boosting, and stacking with other methods [M]. Los Alamitos: IEEE Computer Soc, 2001: 669-670.

[13] 庞 聪,江 勇,廖成旺,等. 基于AdaBoost集成学习的强震动观测抗干扰技术研究[J]. 四川地震, 2020: 14-18.

[14] 赵 刚,黄汉明,卢欣欣,等. 基于BP-Adaboost方法的天然地震和人工爆炸事件波形信号分类识别研究[Z]. 地震工程学报, 2017: 557-563.

[15] ABEYWICKRAMATENINDRA, MUHAMMAD AAMIR CHEEMA, TANIARDAVID. k-Nearest Neighbors on Road Networks: A Journey in Experimentation and In-Memory Implementation [J]. Proceedings of the VLDB Endowment, 2016, 9: 492-503.

[16] WANG P, SHEN C, BARNES N, et al. Fast and robust object detection using asymmetric totally corrective boosting [J]. IEEE Trans Neural Netw Learn Syst, 2012, 23(1): 33-46.

(下转第39页)

- [11] CHO C. Oxygen consumption and denitrification kinetics in soil [J]. *Soil Science Society of America Journal*, 1982,46(4):756-762.
- [12] FILLERY I. Biological denitrification [M]. *Gaseous loss of nitrogen from plant-soil systems*. Springer, 1983:33-64.
- [13] AAMER M, SHAABAN M, HASSAN M U, et al. Biochar mitigates the N₂O emissions from acidic soil by increasing the nosZ and nirK gene abundance and soil pH [J]. *J Environ Manage*, 2020,255.
- [14] WANG X, HU M, REN H, et al. Seasonal variations of nitrous oxide fluxes and soil denitrification rates in subtropical freshwater and brackish tidal marshes of the Min River estuary [J]. *Science of the Total Environment*, 2018,616:1 404-1 413.
- [15] LESCURE C, MENENDEZ L, LENSIS R, et al. Effect of addition of various carbon substrates on denitrification in a vertic Mollisol [J]. *Biology and fertility of soils*, 1992,13(3):125-129.
- [16] ZHAO Y, XIA Y, TI C, et al. Nitrogen removal capacity of the river network in a high nitrogen loading region [J]. *Environ Sci Technol*, 2015,49(3):1 427-1 435.
- [17] GAO J, HOU L, ZHENG Y, et al. nirS-Encoding denitrifier community composition, distribution, and abundance along the coastal wetlands of China [J]. *Applied microbiology and biotechnology*, 2016,100(19):8 573-8 582.
- [18] LIU T, XIA X, LIU S, et al. Acceleration of denitrification in turbid rivers due to denitrification occurring on suspended sediment in oxic waters [J]. *Environ Sci Technol*, 2013,47(9):4 053-4 061.
- [19] ZHE P, ZHANG W-W, SHUAI M, et al. Succession of denitrifying community composition in coastal wetland soils along a salinity gradient [J]. *Pedosphere*, 2012,22(3):367-374.
- [20] 王逢武,刘 玮,万娟娟,等.周丛生物存在下不同水层氧化还原原带的分布及其与微生物的关联[J]. *环境科学*, 2015,36(11):4 043-4 050.
- [21] JORDAN S J, STOFFER J, NESTLERODE J A. Wetlands as sinks for reactive nitrogen at continental and global scales: a meta-analysis [J]. *Ecosystems*, 2011,14(1):144-155.
- [22] SEITZINGER S, HARRISON J A, BÖHLKE J, et al. Denitrification across landscapes and waterscapes: a synthesis [J]. *Ecological applications*, 2006,16(6):2 064-2 090.
- [23] LIVINGSTONE D, PENTECOST A, WHITTON B A. Diel variations in nitrogen and carbon dioxide fixation by the blue-green alga *Rivularia* in an upland stream [J]. *Phycologia*, 2019, 23 (2) : 125-133.

(上接第28页)

- [17] 殷小伟,强志民,贲伟伟,等.污水厂不同生物处理工艺对抗生素的去除效果[J]. *中国给水排水*, 2012,28(22):22-26.
- [18] 汪 峰,钱 庄,张 周,等.污水处理厂尾水对排放河道水质的影响[J]. *安徽农业科学*, 2016,44(14):65-68.
- [19] CHEN K Y, CHEN H X, ZHOU C L, et al. Comparative analysis of surface water quality prediction performance and identification of key water parameters using different machine learning models based on big data[J]. *Water Research*, 2020,171(115 454):1-10.
- [20] TIYASHA, TUNG T M, YASEEN Z M. A survey on river water quality modelling using artificial intelligence models: 2000-2020 [J]. *Journal of Hydrology*, 2020,585:124 670.
- [21] 嵇晓燕,杨 凯,陈亚男,等.基于ARIMA和Prophet的集成学习模型在水质预测中的应用研究[J/OL]. *水资源保护*, 2021:1-11. [2022-03-28]. <http://kns.cnki.net/kcms/detail/32.1356.TV.20211008.1634.022.html>.
- [22] 王凯翔.基于多元回归和神经网络途径的太湖富营养化指标分析与预测[J]. *南京师范大学学报*, 2017,17(2):70-74.
- [23] YANG X Y, LIU Q, LUO X Z, et al. Spatial Regression and Prediction of Water Quality in a Watershed with Complex Pollution Sources [J]. *Scientific Reports*, 2017,7(1):8 318.
- [24] 邓延慧,王正文.湖泊沉积物氮磷赋存转化研究进展[J]. *绿色科技*, 2021,23(6):66-71.
- [25] LU J Y, WANG X M, LIU H Q, et al. Optimizing operation of municipal wastewater treatment plants in China: The remaining barriers and future implications [J]. *Environment International*, 2019, 129:273-278.

(上接第33页)

- [6] 高建文,何圣兵,陈雪初,等.生物滤池/生态砾石床处理含氮污染地表水[J]. *中国给水排水*, 2012,28(1):26-30.
- [7] 高 婷.砾间接触氧化在河水水质净化中的应用[J]. *科学咨询*, 2008(11):54.
- [8] 黄宏斌.都市河川复元:以台湾桃园老街溪为例[J]. *中国防汛抗旱*, 2018,28(10):40-45.
- [9] 葛 俊,胡小贞,庞 燕,等.砾间接触氧化法对白鹤溪低污染水体的净化效果[J]. *环境科学研究*, 2015,28(5):816-822.
- [10] 赵琰鑫,陈 岩,吴颖颖.QUAL2K河流水质模拟模型理论与应用指南[专著][M].北京:气象出版社,2015.
- [11] 周小珍.河流综合水质模型QUAL2K的应用综述[J]. *黑龙江生态工程职业学院学报*, 2019,32(4):22-24.
- [12] PARK S S, YONG S L. A multiconstituent moving segment model for water quality predictions in steep and shallow streams [J]. *Ecological Modelling*, 1996,89(1-3):121-131.
- [13] PELLETIER G J, CHAPRA S C, HUA T. QUAL2Kw - A framework for modeling water quality in streams and rivers using a genetic algorithm for calibration [J]. *Environmental Modelling & Software*, 2006,21(3):419-425.
- [14] YANG M D, SYKES R M, MERRY C J. Estimation of algal biological parameters using water quality modeling and SPOT satellite data [J]. *Ecological Modelling*, 2000.
- [15] 周 华.河流综合水质模型QUAL2K应用研究[J]. *中国水利水电科学研究院学报*, 2010,8(1):71-75.