

技术总结

DOI: 10.19853/j.zgjsps.1000-4602.2024.11.007

基于机器学习的高锰酸钾与有机污染物反应活性预测

朱静怡¹, 黄苑曦¹, 衣启航², 伍洋涛¹, 卜令君¹, 周石庆¹

(1. 水安全保障技术及应用湖南省工程研究中心, 湖南 长沙 410082; 2. 湖南大学设计研究院有限公司, 湖南 长沙 410082)

摘要: 高锰酸钾(KMnO₄)氧化能力强、性质稳定,是自来水厂处理微污染原水和突发污染事件的常见应急氧化药剂,高锰酸钾与有机污染物的反应速率常数(k 值)是衡量其与有机污染物反应活性的重要参数。然而有机污染物种类繁多,单纯依靠试验方法或理论计算获得大量 k 值费时费力且成本高。为此,收集了574种有机污染物的数据,以分子指纹(MFs)和pH作为输入特征,利用4种算法(随机森林RF、神经网络NN、极致梯度提升XGBoost和支持向量机SVM),分别建立了预测 k 值的机器学习(ML)模型,发现4个模型均有良好的稳健性和预测能力,其中RF模型的预测性能最好($R^2_{\text{测试集}}=0.818$, $\text{RMSE}_{\text{测试集}}=0.476$)。采用SHAP方法对RF模型进行了模型解释,发现供电子基团、pH等对 k 值的预测影响最大,表明模型正确学习了 k 与污染物结构之间的关系。同时,计算了RF模型的应用域(AD)以验证其适用范围。该模型可准确、便捷地获得 k 值,有助于了解高锰酸钾与有机污染物的反应活性。

关键词: 高锰酸钾; 有机污染物; 反应速率常数; 机器学习; 随机森林算法; SHAP
中图分类号: TU991 **文献标识码:** A **文章编号:** 1000-4602(2024)11-0041-08

Prediction of Reactivity between Potassium Permanganate and Organic Pollutants Based on Machine Learning

ZHU Jing-yi¹, HUANG Yuan-xi¹, YI Qi-hang², WU Yang-tao¹, BU Ling-jun¹,
ZHOU Shi-qing¹

(1. Hunan Engineering Research Center of Water Security Technology and Application, Changsha 410082, China; 2. Hunan University Design and Research Institute Co. Ltd., Changsha 410082, China)

Abstract: Potassium permanganate (KMnO₄) has strong oxidation ability and stable properties, and is a common emergency oxidation agent for treating micro-polluted raw water and sudden pollution in waterworks. The reaction rate constant (k value) of potassium permanganate to organic pollutants is an important parameter to evaluate its reactivity with organic pollutants. However, there are many kinds of organic pollutants, and it is time-consuming and costly to determine a large number of k values simply by means of experiments or theoretical calculations. This paper collected a total of 574 organic pollutants, and employed four algorithms, such as random forest (RF), neural network (NN), extreme gradient

基金项目: 国家自然科学基金资助项目(52100007)

通信作者: 卜令君 E-mail: lingjunbu@hnu.edu.cn

boosting (XGBoost) and support vector machine (SVM), with molecular fingerprints (MFs) and pH serving as input features to establish machine learning (ML) models for predicting k values. The four models all had good robustness and prediction ability, among which the RF model demonstrated the best prediction performance ($R^2_{\text{test}}=0.818$, $\text{RMSE}_{\text{test}}=0.476$). The SHAP (SHapley Additive exPlanations) method was utilized to analyze the RF model, revealing that electron-donating groups and pH exerted the most significant influence on the prediction of k value. This indicated that the model correctly learned the relationship between k values and the chemical structures of organic pollutants. Simultaneously, the application domain (AD) of the RF model was calculated to validate its suitability. The developed model can accurately and conveniently determine the k values, aiding in understanding the reactivity of potassium permanganate with organic pollutants.

Key words: potassium permanganate; organic pollutant; reaction rate constant; machine learning; random forest (RF) algorithm; SHAP

高锰酸钾(KMnO_4)是一种高效、环保的强氧化剂,是自来水厂处理微污染原水和突发污染事件的常见应急氧化药剂,通常用于解决液氯等常规预氧化药剂无法应对的污染事件^[1]。高锰酸钾在应用中有明显优势:成本低,投加方便,性质稳定,不产生卤代副产物,可氧化去除水中大部分有机物、重金属及藻类等^[2-4]。高锰酸钾等氧化剂与有机污染物之间的反应活性具有极强的结构依赖性,对不同目标污染物的反应活性差异较大^[5-7]。高锰酸钾与有机污染物的二级反应速率常数(k 值)是衡量其反应活性的重要参数,可用于预判水中有机污染物的去除效率,评估高锰酸钾氧化在实际工程中的应用可行性。Laszakovits等^[8]汇总了82种化合物的 k 值,讨论了几种常见官能团对高锰酸钾反应活性的影响,发现高锰酸钾倾向于与烯烃发生亲电加成反应或与酚类发生单电子转移反应,而与醚和伯醇的反应活性很低,比如,在处理藻源臭味事件(土臭素和2-甲基异莰醇)时,高锰酸钾通常无法有效去除目标物质^[9]。因此,预知高锰酸钾与特定有机污染物之间的 k 值,可以实现高锰酸钾在水处理中的合理高效应用。

当前 k 值通常是通过实验法或理论计算法获得。实验法通常使用闪光光解或脉冲辐解法^[10],这两种方法都通过受照样品产生的瞬态产物测得 k 值。随着有机污染物种类和数量的不断增加,通过实验法得到每种有机物的 k 值虽然可行,但会大幅增加研究成本。理论计算法主要是基于密度泛函理论(DFT)的过渡态理论对 k 值进行计算,但此方

法计算复杂、过程繁琐。目前,文献记载的高锰酸钾反应速率常数的数据仍很缺乏,所以,亟需一种方便高效、低成本的方法来获取 k 值。

研究人员尝试建立定量构效关系(QSAR)解决上述问题, QSAR模型将物质的结构与反应活性联系起来(通常认为两者呈线性关系),建立 k 值与某些化学描述符的线性回归方程^[11]。随着物质的结构越来越复杂,结构与反应活性已不满足单一的线性关系,于是基于QSAR模型又发展出了机器学习(ML)等方法。现阶段, ML算法越来越普及^[12-13],研究者输入污染物的结构,通过ML算法预测 k 值。ML算法最显著的优点,一是预测前无需了解反应机制,二是能很好地解决非线性、高维问题。目前,研究者们已经成功建立了用于预测有机污染物与 $\text{HO}\cdot$ ^[14-15]、 O_3 ^[16-17]、 $\text{SO}_4^{\cdot-}$ ^[18-19]、 Fe(VI) ^[12,20]、 HClO ^[21]等反应速率常数的ML模型,而用于预测有机污染物与 KMnO_4 反应活性的ML模型尚未见报道。

鉴于此,笔者从文献中收集了574种有机污染物与 KMnO_4 的 k 值,汇总分析所有数据;将分子指纹(MFs)和pH作为输入,对比了4种机器学习算法(随机森林RF、神经网络NN、极致梯度提升XGBoost和支持向量机SVM)的预测效果,得到了最佳的ML模型,并预测了有机污染物与 KMnO_4 的 k 值;采用SHAP(SHapley Additive exPlanations)分析模型的预测结果,探讨了分子结构对 k 值的影响;最后对模型进行了应用域(AD)分析。

1 材料与方法

本研究分为6个阶段,流程如图1所示。

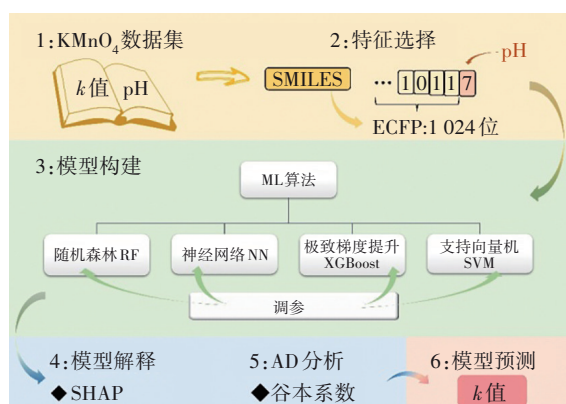


图1 研究流程

Fig.1 Workflow of this study

1.1 数据集

根据已有文献^[2,8,22-24]汇总得到含有574种有机污染物的数据集,该数据集包含了有机污染物与高锰酸钾的 k 值和实验条件(温度和pH)。在收集数据时,仅收集实验测得的有机物污染物的 k 值;根据已有文献和化学知识,排除不合理的记录;另外,排除实验条件在低温($<16\text{ }^{\circ}\text{C}$)和高温($>35\text{ }^{\circ}\text{C}$)条件下的数据。如果同一化合物在不同文献中有多个 k 值,则取平均值。经过筛选,数据集一共包括574个数据点。所有 k 值经过对数转换为 $\lg k$, $\lg k$ 的范围为 $-4.510\ 0\sim 4.322\ 2$ 。将574个数据以4:1的比例随机分为训练集和测试集,训练集包含459个数据,测试集包含115个数据。所有代码采用Python(3.9.9版本)软件编写。

1.2 特征选择

本研究中模型的分子结构描述符选用分子指纹(MFs)中有1 024个特征位点的ECFP。ECFP是由“0”和“1”组成的数组,其中,“0”代表物质中不含该特征,“1”代表物质中含有该特征,每个特征都代表一种分子结构。ECFP的原理是将分子中的每个原子依次看作是一个环形(即分子子结构)的中心,为每个原子分配一个唯一的编号,利用原子之间的连接关系,通过迭代形成唯一的指纹。物质的活性与分子结构密切相关,ECFP包含了丰富的分子结构信息,因此可以帮助预测物质的活性。

首先,从PubChem网站上搜集574种有机污染物的SMILES(simplified molecular input line entry system)。在Python软件中采用RDkit安装包将SMILES统一转化成MFs。然后通过Python软件 pandas 筛选描述符,删除与分子活性相关性低的特

征,建模采用的描述符应满足该位点所有值的方差大于0.3。经计算,本研究在1 024个特征中选择了210个特征。

pH会影响可电离化合物的 k 值^[20-21],然而目前构建ML模型时很少考虑pH,这可能会导致一个化合物具有多个不同的 k 值,误导模型的学习过程;同时,忽略pH将会导致模型对数据集的普适性较差和数据量减少的问题。因此,本研究将反应的pH作为输入特征,与ECFP一并组成特征数组。故本研究共筛选出211个输入特征。

1.3 模型的构建和评价

在Python软件中采用由scikit-learn开发的4种机器学习算法(RF、NN、XGBoost和SVM)构建模型,将MFs、pH和对应的 k 值输入训练模型。训练集被分成5个相等的部分,进行5倍交叉验证。每种算法的原始参数可以用于训练模型,为了使算法达到最佳性能,通常研究者会调整超参数,本研究使用Python中的随机搜索调整各算法的超参数。模型的性能一般通过均方根误差(RMSE)和相关系数(R^2)等指标来评估, RMSE越低、 R^2 越高,模型的效果越好。根据测试集的RMSE和 R^2 值,筛选出最佳的机器学习算法及其超参数。最后,通过采用最优算法及其最优超参数的模型对所有数据集进行重新训练,得到最优模型。

1.4 模型解释

ML模型不仅要符合统计规律,还应与基本化学原理一致。SHAP(SHapley Additive exPlanations)方法是一种模型解释包,可计算每个特征对预测的贡献,并对其进行可视化和排序。本研究采用SHAP方法来分析211个特征(化合物的结构和pH)与 k 值之间的内在联系,直观地了解对 k 值影响最大的输入特征,并与常规化学原理进行对比。

1.5 应用域(AD)分析

应用域(AD)分析可以验证模型预测的训练集化合物的 k 值是否可靠^[25]。本研究采用谷本系数 $T(A,B)$ 作为AD参数。 $T(A,B)$ 指测试集化合物和每个训练集化合物之间的相似度,计算方法^[26]如下:

$$T(A,B) = \frac{N_c}{N_a + N_b + N_c} \quad (1)$$

式中:A、B为两种不同的化合物; N_a 、 N_b 分别为化合物A和B的MFs中“1”的数量,即物质中包含特

征的数量; N_c 为化合物A和B的MFs均为“1”的数量,即两种物质共同包含的特征的数量。

当一个化合物位于模型的AD中时,则可认为QSAR模型对该化合物的预测结果是可信的。本研究将训练集化合物的平均相似度值用作相似度指标。通过比较每个测试集数据对所有训练集数据的平均相似度,假设 $T(A,B)$ 阈值。如果计算得到的 $T(A,B)$ 高于预先设定的阈值,那么该化合物在AD范围内,此时模型对该化合物的预测是可靠的、稳健的;若计算出的 $T(A,B)$ 低于预先设定的阈值,则从测试集中删去该化合物,模型对其的预测结果是不可靠的。对于任何模型来说,删去的化合物越少,则它的适用范围越广。

2 结果与分析

2.1 Meta分析

建模之前,首先对所有收集到的数据进行统计分析。根据基团和反应活性,将所有化合物分为烯烃、苯酚、烷基苯、醇、醚、酮、卤化物和硝基化合物八类。当一个化合物有多个基团时,该化合物将根据各基团被分配到多个类别,各类 $\lg k$ 的箱型图如图2所示。总的来看, $\lg k$ 从最低的-4.5100(二噁烷)到最高的4.3222(鱼腥藻毒素a),差了8.8322个数量级,其中硝基化合物的 $\lg k$ 范围最广,达到了7.4523个数量级,说明 KMnO_4 对污染物的反应活性差异大,具有很高的选择性。含有供电子基团的各类物质的 $\lg k$ 的中位数(烯烃、苯酚、烷基苯、醇和醚的 $\lg k$ 的中位数依次为1.7634、1.2245、2.0470、1.4552和1.4552)大体上高于含有吸电子基团的物质(酮、卤化物和硝基化合物的 $\lg k$ 的中位数依次为2.2178、0.9243和-1.0092),这说明供电子基团加快了反应速率,吸电子基团则降低了反应速率^[8]。但同时

也发现,酮类的 $\lg k$ 的中位数是所有类别中最高的,出现这种反常现象是因为收集的大部分酮类化合物含有其他供电子基团, KMnO_4 会先与其他反应位点反应,导致酮类的 $\lg k$ 升高。可见,单纯的统计学手段无法有效获取 $\lg k$ 与物质结构之间的本质关系。

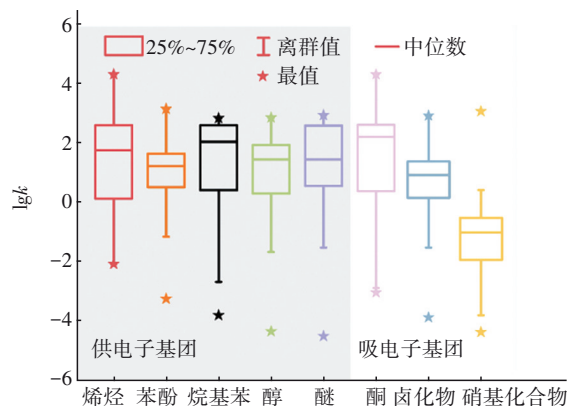


图2 各类有机污染物的 $\lg k$ 的箱型图

Fig.2 Box plot of $\lg k$ values of various organic pollutants

2.2 ML模型的预测效果

本研究建立的4个模型的预测效果见表1。对于训练集,SVM、RF、XGBoost和NN模型的 R^2 分别达到了0.953、0.942、0.894和0.888, RMSE分别为0.333、0.368、0.497和0.511;对于测试集,4个模型的预测性能虽然有所下降,但是仍保持在较高水平, R^2 分别为0.782、0.818、0.802、0.754, RMSE分别为0.520、0.476、0.496、0.553。较高的 R^2 和较低的RMSE说明4个模型对高锰酸钾与有机污染物的反应活性都具有较好的预测性能;同时,较低的 ΔR^2 ($|R^2_{\text{测试集}} - R^2_{\text{训练集}}|$)和 ΔRMSE ($|\text{RMSE}_{\text{测试集}} - \text{RMSE}_{\text{训练集}}|$)说明了4个模型没有过度拟合。综上,4个模型均具有良好的预测能力、拟合度和统计稳健性。

表1 模型预测效果

Tab.1 Performance of the ML models

ML算法	训练集数据量	测试集数据量	$R^2_{\text{训练集}}$	$R^2_{\text{测试集}}$	$\text{RMSE}_{\text{训练集}}$	$\text{RMSE}_{\text{测试集}}$	ΔR^2	ΔRMSE
RF	459	115	0.942	0.818	0.368	0.476	0.124	0.108
NN			0.888	0.754	0.511	0.553	0.134	0.042
XGBoost			0.894	0.802	0.497	0.496	0.092	0.001
SVM			0.953	0.782	0.333	0.520	0.171	0.187

本研究以测试集的 R^2 和RMSE作为分析指标来比较各模型性能。如表1所示,RF模型的 R^2 略高于NN、XGBoost和SVM模型,其RMSE则略低于另外3个模型,说明在4个模型中,RF模型的预测效果最

好。图3是各模型的 $\lg k$ 实测值与预测值的散点图,红点和蓝点越靠近虚线($y=x$)说明预测误差越小,拟合度越高。可以看出,RF模型的散点分布更接近虚线($y=x$),说明RF模型预测更准确。综上可知,RF

算法更适用于预测 $\lg k$ 。

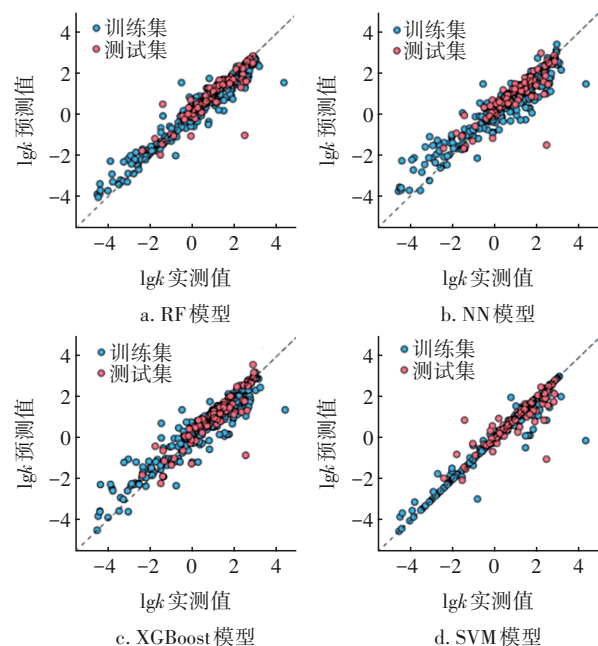


图3 4种模型的 $\lg k$ 实测值与预测值

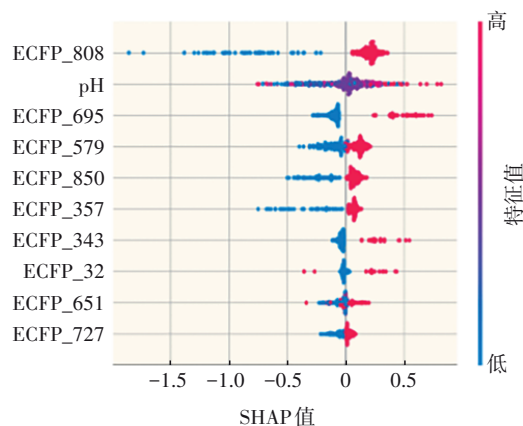
Fig.3 Experimental and predictive $\lg k$ of four models

根据表1和图3可知,模型预测性能的排名为 $\text{RF} > \text{XGBoost} > \text{SVM} > \text{NN}$ 。RF和XGBoost是两种常见的树算法,树算法擅长处理高维度的复杂数据集同时避免过度拟合,能很好地学习稀疏数据的规律。在以往的研究中树算法都有不错的表现,Zhong等人^[21]利用树算法成功建立了预测 O_3 、 SO_4^{2-} 、 HClO 和 ClO_2 的统一学习和迁移学习模型。本研究建立的4个模型的预测性能排名再一次证明了树算法在物质反应活性预测领域的实用性和可靠性。

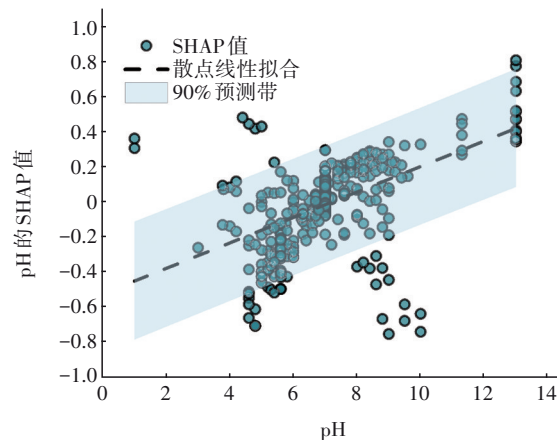
2.3 模型解释

为了验证模型预测是否科学,确定输入特征对模型预测性能的重要性,本研究对RF模型进行了SHAP分析,结果见图4和表2。图4(a)是RF模型的SHAP图,y轴是对预测结果影响排名前十位的特征,y轴上的数字,如808、695和579,表示某一官能团在ECFP中的位置,pH指各文献记载的实验法测得有机污染物的 k 值时的pH,x轴是SHAP值的正负,图中每个点代表一个化合物的SHAP值。红点(“1”)表示该化合物中存在y轴上的官能团,而蓝点(“0”)表示有机污染物没有该官能团,正值表示该官能团对预测的 k 值有正面影响,负值表示该官能团对预测的 k 值有负面影响。表2是图4(a)中对 k 值预测的影响排在前十的特征输入,蓝点代表中心

原子,未标记的蓝点代表碳原子,黄点代表芳香原子,黑色实线代表特征中的键,灰线代表不在特征中的相邻键,虚线代表共轭结构。



a. RF模型的SHAP图



b. pH的部分依赖图

图4 RF模型的SHAP分析结果

Fig.4 SHAP interpretation of RF model

表2 影响排在前十的结构

Tab.2 The top 10 influential functional groups

排名	特征	影响	排名	特征	影响
1		ECFP_808增强	6		ECFP_357增强
2	pH		7		ECFP_343增强
3		ECFP_695增强	8		ECFP_32增强
4		ECFP_579增强	9		ECFP_651增强
5		ECFP_850增强	10		ECFP_727增强

如图4(a)和表2所示,对有机污染物的 k 值最重要的10个特征分别为 ECFP_808 ($>\text{C}=\text{C}$)、pH、

ECFP_695($-\text{C}=\text{C}-$)、ECFP_579($\text{C}_6\text{H}_5-\text{OH}$)、ECFP_850(共轭结构)、ECFP_357(共轭结构)、ECFP_343($=\text{C}$)、ECFP_32($-\text{C}=\text{C}-\text{C}=\text{C}-$)、ECFP_651($=\text{O}$)和ECFP_727($\text{C}_6\text{H}_5-\text{C}$)。可见,分子结构中碳碳双键(ECFP_808、ECFP_695、ECFP_343、ECFP_32)对高锰酸钾反应活性是最重要的,起到增强的作用;酚羟基(ECFP_579)也能提高 k 值,其作用仅次于碳碳双键;两个共轭结构(ECFP_850、ECFP_357)以及烷基苯(ECFP_727)也起到了促进作用,上述结果均说明供电子基团会提高 KMnO_4 与有机污染物的 k 值。这与文献报道一致,即高锰酸钾与有机污染物反应主要通过亲电攻击双键、与酚类物质的单电子转移、与碳氢键的抽氢反应等3条路径开展^[8,27],因此,尽管在训练模型时未考虑官能团对 k 值的影响,模型仍正确学习了官能团与 k 值之间的关系,这进一步证实了所建模型从数理和机理两方面都具有较强的可行性。

如图4(a)所示,虽然pH对高锰酸钾 k 值的影响在模型运行时排在第2位,但pH的影响不同于分子指纹片段的影响机制清晰,因此,进一步绘制了pH的部分依赖图,如图4(b)所示。可以看出,pH对有机污染物的 k 值影响较为复杂,整体而言,超过90%的物质落在蓝色区域内,说明随着pH的升高,有机污染物的 k 值呈现上升趋势,该现象可从两方面解释^[28-29]:①当 $\text{pH}>9.0$ 时存在高锰酸钾的碱催化现象;②由于酚酸盐的电子云密度远高于酚类物质,故pH在5.0~9.0范围内时酚酸盐比酚类物质更容易发生氧化反应, k 值也更高。但也有30多个点离散在蓝色区域以外(集中在图中左上和右下区域),这是由于:①当 $\text{pH}<5.0$ 时存在高锰酸钾的酸催化现象;②部分物质(如烯烃类物质)在pH为5.0~10.0范围内不发生电离,故pH对此类物质 k 值的影响可忽略不计^[30]。

综上所述,本研究建立的ML模型可以做出符合化学基本原理的预测,有助于揭示污染物化学结构与 k 值之间的内在联系。

2.4 应用域(AD)分析

RF模型的AD分析通过Python软件计算谷本系数 $T(\text{A}, \text{B})$ 来实现,结果如表3所示,随着阈值从0.02增至0.06,AD以外的有机污染物数量从0增加到9,计算出的 $\text{RMSE}_{\text{测试集}}$ 也在增加。显然,阈值为0.02和0.04时, $\text{RMSE}_{\text{测试集}}$ 最小为0.476,但考虑

到测试集删去的化合物应尽可能少,故将阈值定为0.02。因此,如果化合物与训练数据集的平均结构相似度超过0.02,则认为RF模型可以对该污染物与 KMnO_4 的 k 值做出正确的预测。阈值设置越低,则该模型越稳健,适用范围越广,再一次证明了构建的RF模型具有良好的实用性和稳健性。

表3 RF模型的AD分析结果

Tab.3 Application domain of RF model

数据量	最佳模型	$\text{RMSE}_{\text{测试集}}$	阈值	AD外的污染物数量	计算的 $\text{RMSE}_{\text{测试集}}$
574	RF模型	0.476	0.06	9	0.486
			0.04	1	0.476
			0.02	0	0.476

3 结论

① RF模型对 KMnO_4 与有机污染物反应活性的预测准确($R^2_{\text{测试集}}=0.818$ 、 $\text{RMSE}_{\text{测试集}}=0.476$)、适用范围广。

② KMnO_4 反应具有选择性,更倾向于与富电子官能团反应,供电子基团的存在会增加有机污染物的 k 值,吸电子基团则会降低 k 值。

③ 本研究中大部分有机污染物表现出了明显的pH依赖性,即pH越高, k 值越高,但也有部分不可电离有机污染物不遵循此规律。

④ ML模型依赖于输入的数据,选择的训练数据会影响模型的表现,从而导致模型在学习输入特征与输出值的关系时出现误差。因此,数据的质量是能否成功建模的先决条件之一,然而,数据的稀缺是水环境研究最常面临的问题之一,许多非共享的数据集、训练模型或源代码也增加了研究难度。目前,研究人员也尝试建立了一些大型共享数据集,如:包含100多万种材料的量子计算结果的OQMD数据集,但如何维护、继续开发这些公开数据库仍是研究者需要考虑的问题。

⑤ ML模型可以高效准确地预测反应速率常数,除此之外,ML还能分析解决其他水环境问题,如:饮用水和废水处理、能源消耗和污染物理化性质预测等,但这并不意味着ML可以完全取代实验,因为任何ML模型都建立在实验得到的历史数据之上。因此,在未来,研究者不能忽略实验方法,研究者应将实验研究和机器学习有机结合,共同解决环境问题,例如,通过实验来产生数据,通过ML发掘实验中的化学知识。

参考文献:

- [1] 邓俊杰, 劳志雄, 陈明良, 等. 强化老旧水厂常规处理工艺优化提升出水水质[J]. 中国给水排水, 2021, 37(2): 85-89.
- DENG Junjie, LAO Zhixiong, CHEN Mingliang, *et al.* Optimizing and improving effluent quality by strengthening conventional water treatment process in an aging waterworks [J]. China Water & Wastewater, 2021, 37(2): 85-89 (in Chinese).
- [2] CHEN J, LING J, SUN B, *et al.* Trace organic contaminants abatement by permanganate/bisulfite pretreatment coupled with conventional water treatment processes: lab- and pilot-scale tests [J]. Journal of Hazardous Materials, 2021, 401: 123380.
- [3] CAO W, WU N, QU R, *et al.* Oxidation of benzophenone-3 in aqueous solution by potassium permanganate: kinetics, degradation products, reaction pathways, and toxicity assessment [J]. Environmental Science and Pollution Research, 2021, 28(24): 31301-31311.
- [4] GUAN X, HE D, MA J, *et al.* Application of permanganate in the oxidation of micropollutants: a mini review [J]. Frontiers of Environmental Science & Engineering in China, 2010, 4(4): 405-413.
- [5] JIANG Y, GOODWILL J E, TOBIASON J E, *et al.* Bromide oxidation by ferrate (VI): the formation of active bromine and bromate [J]. Water Research, 2016, 96: 188-197.
- [6] TUFAIL A, PRICE W E, MOHSENI M, *et al.* A critical review of advanced oxidation processes for emerging trace organic contaminant degradation: mechanisms, factors, degradation products, and effluent toxicity [J]. Journal of Water Process Engineering, 2021, 40: 101778.
- [7] 刘臻, 昌子雄, 刘浪, 等. 基于高锰酸钾的高级氧化技术降解水中新兴污染物研究综述[J]. 水处理技术, 2023, 49(2): 18-24.
- LIU Zhen, CHANG Zixiong, LIU Lang, *et al.* Review of advanced oxidation technology based on potassium permanganate for the degradation of emerging pollutants in water [J]. Technology of Water Treatment, 2023, 49(2): 18-24 (in Chinese).
- [8] LASZAKOVITS J R, KERR A, MACKAY A A. Permanganate oxidation of organic contaminants and model compounds [J]. Environmental Science & Technology, 2022, 56(8): 4728-4748.
- [9] 杜崇岭, 秦新玲, 王令刚, 等. 季节性原水水质变化除臭应用实践[J]. 给水排水, 2022, 48(S1): 577-581.
- DU Chongling, QIN Xinling, WANG Linggang, *et al.* Deodorization application practice of seasonal source water quality change [J]. Water & Wastewater Engineering, 2022, 48(S1): 577-581 (in Chinese).
- [10] SHI Y, LI J J, WANG Q, *et al.* Computer-aided estimation of kinetic rate constant for degradation of volatile organic compounds by hydroxyl radical: an improved model using quantum chemical and norm descriptors [J]. Chemical Engineering Science, 2022, 248: 117244.
- [11] 张坤, 赵健辰, 金文杰, 等. 水相中有机化学品和卤素自由基反应速率常数的定量结构-活性关系模型[J]. 环境化学, 2022, 41(2): 663-672.
- ZHANG Kun, ZHAO Jianchen, JIN Wenjie, *et al.* A quantitative structure-activity relationship model for reaction rate constants of organic chemicals with halogen radicals in aqueous phase [J]. Environmental Chemistry, 2022, 41(2): 663-672 (in Chinese).
- [12] GAO Y, ZHONG S, ZHANG K, *et al.* Abiotic reduction of organic and inorganic compounds by Fe(II)-associated reductants: comprehensive data sets and machine learning modeling [J]. Environmental Science & Technology, 2023, 57(46): 18026-18037.
- [13] HOUSTON P L, NANDI A, BOWMAN J M. A machine learning approach for rate constants. III. Application to the $\text{Cl}(\text{P}) + \text{CH}_4 \rightarrow \text{CH}_3 + \text{HCl}$ reaction [J]. The Journal of Physical Chemistry A, 2022, 126(33): 5672-5679.
- [14] ZHONG S, HU J, FAN X, *et al.* A deep neural network combined with molecular fingerprints (DNN-MF) to develop predictive models for hydroxyl radical rate constants of water contaminants [J]. Journal of Hazardous Materials, 2020, 383: 121141.
- [15] LUO X, WEI X, CHEN J, *et al.* Rate constants of hydroxyl radicals reaction with different dissociation species of fluoroquinolones and sulfonamides: combined experimental and QSAR studies [J]. Water Research, 2019, 166: 115083.
- [16] CHENG Z, YANG B, CHEN Q, *et al.* A quantitative-structure-activity-relationship (QSAR) model for the reaction rate constants of organic compounds during the ozonation process at different temperatures [J]. Chemical Engineering Journal, 2018, 353: 288-296.
- [17] ZHU H, GUO W, SHEN Z, *et al.* QSAR models for degradation of organic pollutants in ozonation process

- under acidic condition [J]. *Chemosphere*, 2015, 119: 65–71.
- [18] YE T, WEI Z, SPINNEY R, *et al.* Chemical structure-based predictive model for the oxidation of trace organic contaminants by sulfate radical [J]. *Water Research*, 2017, 116: 106–115.
- [19] XIAO R, YE T, WEI Z, *et al.* Quantitative structure-activity relationship (QSAR) for the oxidation of trace organic contaminants by sulfate radical [J]. *Environmental Science & Technology*, 2015, 49(22): 13394–13402.
- [20] ZHENG S S, GUO W Q, LU H, *et al.* Machine learning approaches to predict the apparent rate constants for aqueous organic compounds by ferrate [J]. *Journal of Environmental Management*, 2023, 329: 116904.
- [21] ZHONG S, ZHANG Y, ZHANG H. Machine learning-assisted QSAR models on contaminant reactivity toward four oxidants: combining small data sets and knowledge transfer [J]. *Environmental Science & Technology*, 2022, 56(1): 681–692.
- [22] MAHMOODLU M G, HASSANIZADEH S M, HARTOG N. Evaluation of the kinetic oxidation of aqueous volatile organic compounds by permanganate [J]. *Science of the Total Environment*, 2014, 485/486: 755–763.
- [23] JIANG J, PANG S Y, MA J. Role of ligands in permanganate oxidation of organics [J]. *Environmental Science & Technology*, 2010, 44(11): 4270–4275.
- [24] ZHANG Y, SUN B, RAO D, *et al.* Could manganate be an alternative of permanganate for micropollutant abatement? [J]. *Chemosphere*, 2023, 321: 138094.
- [25] SANCHES-NETO F O, DIAS-SILVA J R, JUNIOR L H K Q, *et al.* “pySiRC”: machine learning combined with molecular fingerprints to predict the reaction rate constant of the radical-based oxidation processes of aqueous organic contaminants [J]. *Environmental Science & Technology*, 2021, 55(18): 12437–12448.
- [26] ZHONG S, ZHANG K, WANG D, *et al.* Shedding light on “black box” machine learning models for predicting the reactivity of HO radicals toward organic compounds [J]. *Chemical Engineering Journal*, 2021, 405: 126627.
- [27] 许可, 贾伟伟, 强志民. 高锰酸钾氧化降解水中微量有机污染物的研究进展[J]. *环境化学*, 2017, 36(1): 16–26.
- XU Ke, BEN Weiwei, QIANG Zhimin. Oxidation degradation of micro-organic pollutants by potassium permanganate in water: a review [J]. *Environmental Chemistry*, 2017, 36(1): 16–26 (in Chinese).
- [28] WIBERG K B, FREEMAN F. Kinetics of the base-catalyzed permanganate oxidation of benzaldehyde [J]. *The Journal of Organic Chemistry*, 2000, 65(2): 573–576.
- [29] BANO F, STEWART R. Permanganate oxidation of aromatic alcohols in acid solution [J]. *Canadian Journal of Chemistry*, 2011, 47: 3199–3205.
- [30] LI J, PANG S Y, WANG Z, *et al.* Oxidative transformation of emerging organic contaminants by aqueous permanganate: kinetics, products, toxicity changes, and effects of manganese products [J]. *Water Research*, 2021, 203: 117513.

作者简介: 朱静怡(2000–), 女, 湖南长沙人, 硕士研究生, 研究方向为机器学习在环境相关问题中的应用。

E-mail: 1004698998@qq.com

收稿日期: 2023-08-29

修回日期: 2023-10-12

(编辑: 刘贵春)