

基于无人机影像和机器学习的土壤铬和锌预测

张帅, 韩英越, 代聪, 刘凯, 吕运鸿, 彭靖宇

(浙江大学 建筑工程学院, 浙江 杭州 310058)

摘要: 本文以安徽某重金属污染场地的地表土壤为研究对象, 采集236个土壤样本, 测定了土壤重金属Cr和Zn的含量。进行无人机航测获得场地影像及地形模型, 提取影像中的地形因子及采样点与周边潜在污染源的距離因子作为预测模型自变量。采用梯度提升决策树(Gradient Boosting Decision Tree, GBDT)、随机森林(Random Forest, RF)、极端梯度提升(Extreme Gradient Boosting, XGBoost)3种机器学习回归算法进行土壤Cr和Zn含量的预测。结果表明, 场地地形因子与地表土壤重金属的分布存在相关性。3种机器学习算法对土壤重金属的预测有很好的效果, 其中GBDT对土壤Cr和Zn含量的预测效果最好, 对应的平均绝对误差(MAE)分别为12.4 mg/kg和13.6 mg/kg, 对应的均方根误差(RMSE)分别为17.6 mg/kg和18.8 mg/kg。GBDT相对于克里金插值法对于Cr和Zn预测的MAE分别降低了33.6%和52.1%, RMSE分别降低了29.6%和51.2%。在选取的特征因子中, 工厂距离对于Cr和Zn含量的预测贡献度较大, 贡献度分别为15.5%和15.1%。本研究提出的基于无人机影像和机器学习的污染识别框架对于污染场地重金属含量的预测提供了新的思路, 对污染场地的风险识别及修复治理有重要意义。

关键词: 土壤; 污染; 重金属; 无人机; 影像; 机器学习

中图分类号: TU195

文献标识码: A

文章编号: 2096-7195(2024)02-0125-09

Prediction of chromium and zinc contents in soil based on UAV imagery and machine learning

ZHANG Shuai, HAN Ying-yue, DAI Cong, LIU Kai, LV Yun-hong, PENG Jing-yu

(College of Civil Engineering and Architecture, Zhejiang University, Hangzhou 310058, Zhejiang, China)

Abstract: In this study, the soil of a heavy metal contaminated site in Anhui Province was used as the research object, and the contents of heavy metals Cr and Zn of 236 soil samples were obtained. The UAV aerial survey was conducted to obtain the site image and terrain model. Then the terrain factor in the image and the distance factor between the sampling point and the surrounding potential pollution sources were extracted as the independent variables of the prediction model. Three machine learning regression algorithms, Gradient Boosting Decision Tree (GBDT), Random Forest (RF), and Extreme Gradient Boosting (XGBoost), were used for the prediction of soil Cr and Zn contents. The results show that there was a correlation between terrain factors and the distribution of heavy metals in the site. The three machine learning algorithms have good effect on the predictions of soil heavy metals, among which GBDT has the best prediction effect on soil Cr and Zn content, corresponding to a mean absolute error (MAE) of 12.4 mg/kg and 13.6 mg/kg, and a root mean square error (RMSE) of 17.6 mg/kg and 18.8 mg/kg. GBDT reduces the MAE by 33.6% and 52.1%, and RMSE by 29.6% and 51.2% compared to kriging interpolation for Cr and Zn prediction. Among the selected characteristic factors, the contribution of factory distance to the prediction of Cr and Zn contents is greater, with a contribution of 15.5% and 15.1%, respectively. The contamination identification framework based on UAV images and machine learning proposed in this study provides a new idea for the prediction of heavy metal content in contaminated sites, which is important for the risk identification and remediation management of contaminated sites.

Key words: soil; pollution; heavy metal; UAV; imagery; machine learning

收稿日期: 2022-10-18

基金项目: 国家自然科学基金(52278376)。

作者简介: 张帅(1985—), 女, 浙江杭州人, 博士, 研究员, 主要从事岩土工程方面相关工作。E-mail: zhangshuaiqj@zju.edu.cn。

0 引言

重金属一般是指密度大于 4.5 g/cm^3 的金属和类金属,如铅(Pb)、砷(As)、铜(Cu)、镍(Ni)、铬(Cr)、锰(Mn)、锑(Sb)、镉(Cd)、汞(Hg)、锌(Zn)等^[1]。随着农业、城市化、工业化、采矿等人类活动的发展,土壤介质受到的重金属污染越来越严重^[2]。土壤中重金属的人为来源主要是固体废物的填埋、工业生产、运输、采矿和冶炼、污水灌溉、化肥和农药的使用^[3]。土壤重金属污染目前是一个全球性的威胁,在世界许多地区都很严重^[4]。重金属具有持久性和生物蓄积的特点,土壤中重金属随着时间的推移在作物中积累^[5-6],最终可通过摄入、吸入和皮肤接触被人体吸收,它们通过循环系统到达特定的人体组织,从而造成健康危害^[7-8]。

确定污染场地的土壤重金属含量的分布对于污染场地的风险评估及修复治理十分重要。但污染场地重金属的分布广泛且分散,且物理化学运移的不确定性非常高^[9]。当前污染场地土壤重金属的分布情况普遍通过密集的网络布点,然后进行现场取样及室内重金属含量测试^[10]。对于未采样的土壤点位,常采用克里金等插值方法进行重金属含量的预测^[11]。传统插值方法主要基于空间自相关,即假设近观测值的属性值比远观测值的相关性更强^[12]。总体上,此方法耗时费力。

因此开展高效低成本的土壤重金属含量识别预测技术研究是十分有必要的。近年来,一些学者尝试通过将遥感与地球化学技术结合进行污染场地土壤重金属分布研究,如利用卫星^[13-14]或者无人机^[15-16]。

基于卫星或无人机获取的影像可以提取影响土壤重金属分布的环境特征,进而可以采用机器学习的方法处理大量复杂特征因子与土壤重金属含量的关系。李国旭等^[17]利用 WorldView-3 卫星光谱影像以及相关的环境变量,使用随机森林、支持向量机、偏最小二乘回归、BP 神经网络这 4 种回归算法对某流域农田中 Cd 的含量进行了识别预测,并证明了相关环境变量是进行土壤重金属含量预测的最重要特征。尹芳等^[18]利用分段偏最小二乘法处理 Sentinel-2 卫星影像,建立了土壤中 Cu 含量与光谱特征的相关关系,并进行了土壤中 Cu 含量的预测。王腾军等^[19]结合 Landsat 8 卫星影像光谱数据特征及坡度特征,利用随机森林算法

进行了某矿区土壤 As、Cu 和 Pb 含量的预测。这些研究都基于卫星影像进行污染场地土壤重金属含量的预测。相比于卫星影像,无人机具有图像采集时间短、数据分辨率高等优势^[20]。目前利用无人机拍摄高精度污染场地影像,充分提取无人机影像中的环境变量,进行环境变量与土壤重金属含量的相关性分析,进而进行土壤重金属含量预测的研究少有报道。

因此本文以安徽某重金属污染场地为研究对象,利用无人机获取场地影像,提取影响地表土壤 Cr 和 Zn 含量的特征因子,采用梯度提升决策树(GBDT)、随机森林(RF)、极端梯度提升(XGBoost) 3 种机器学习回归算法进行土壤重金属含量的预测。研究结果可为探索基于无人机影像和机器学习算法进行污染场地土壤重金属含量的预测评估提供新的思路。

1 研究区概况

研究区位于安徽省某污染场地(图 1),研究区属亚热带湿润季风气候,年降雨量在 $1\ 200 \sim 1\ 500 \text{ mm}$ 之间。采样区总面积约为 $33\ 000 \text{ m}^2$ 。研究区域东南方向有 1 个工厂、西北方向有 1 个自来水厂、研究区南部有 1 个村庄。同时场地被 1 条主要公路、1 条河流和 1 条铁路所包围。这些在本研究中被认为是土壤污染的潜在污染源。受工业活动、交通及其他人为活动的影响,该场地地表土壤受到重金属的污染。



图 1 研究区概况

Fig. 1 Overview of the study area

2 数据与方法

基于无人机影像和机器学习算法的土壤中重金属含量预测的研究技术路线如图 2 所示。进行现场土壤点位的布设及样品的采集,获得土壤重金属 Cr 和 Zn 的含量,作为本次预测模型的训练及测

试的原始数据。进行场地无人机影像的拍摄, 并获得场地的影像及数字高程模型 (DEM), 进而从无人机影像中获取场地的地形数据及采样点与

潜在污染源的的距离作为影响土壤重金属分布的特征因子。最后基于 GBDT、RF、XGBoost 这 3 种机器学习回归算法进行土壤重金属含量的预测。

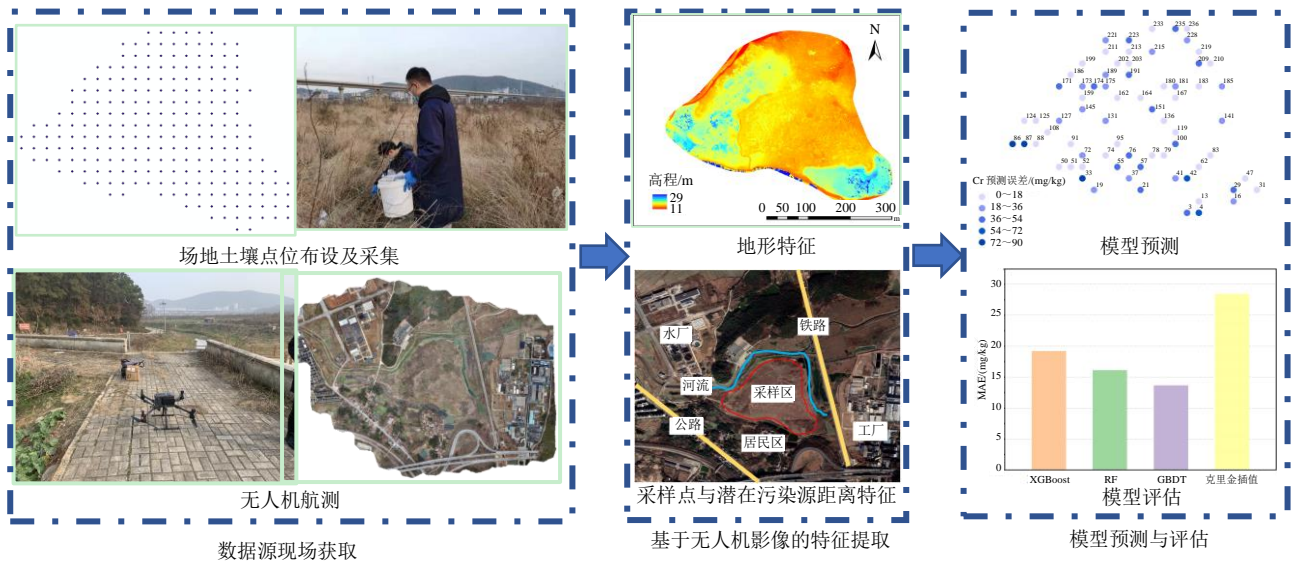


图 2 本文研究技术路线

Fig. 2 Technical route of the research in this paper

2.1 土壤样品的采集及测试

采样点按照 25 m 等间距网格为基础进行布设 (图 3)。研究区利用 GPS 进行采样点的定位, 共采集表层土壤 (0~20 cm) 236 个样本。在每个采样点位置处, 将 5 个表层土壤样本组合为该采样点的代表样本。所有样品都用木铲采集, 并储存在玻璃瓶中。土壤样本在室内风干, 并去除杂质和石头等。然后, 土壤样品在磨粉机中研磨, 并通过 0.15 mm 的尼龙筛进行筛分。土壤中 Cr 和 Zn 的含量根据《土壤和沉积物铜、锌、铅、镍、铬的测定火焰原子吸收分光光度法》(HJ 491—2019) 测定。

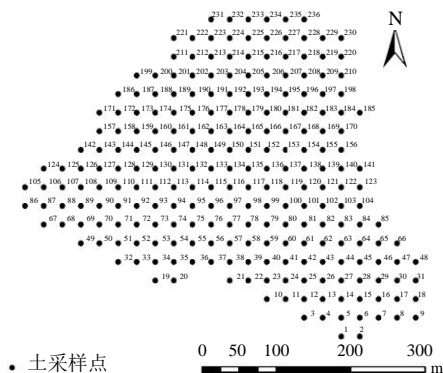


图 3 土壤采样点布设图

Fig. 3 Layout map of soil sampling points

2.2 无人机影像的采集

场地的无人机影像由大疆 M300 RTK 四轴飞行器采集。由于该无人机多旋翼的特点, 可以稳定和安全的飞行。本次无人机的平均飞行速度为 6 m/s,

飞行高度距地面 100 m, 航向重叠 80%, 侧面重叠 70%。获得的无人机影像分辨率为 5 cm。

2.3 机器学习回归算法

本研究采用 GBDT、RF、XGBoost 这 3 种机器学习回归算法进行土壤重金属含量的预测。在 236 个样本点中, 随机选取 70% 的样本点作为训练数据集, 剩余 30% 作为验证数据集。GBDT 是一种基于回归决策树的迭代优化算法, 与传统回归模型相比, GBDT 模型可以提供更准确的预测结果和更高的计算效率^[21]。此外, GBDT 模型采用梯度提升算法来强调接近正确预测的数据, 因此对缺失数据和异常值具有鲁棒性^[22]。RF 是一种基于分类回归树和 bagging 学习策略的预测建模算法^[23], 每个随机选择的子样本及其对应的属性被用来生成一个分类树或回归树, 所有的树形成一个森林, 用所有树预测的平均值作为回归结果。XGBoost 是一种高效且可扩展的梯度增强算法实现^[24], 该算法的主要特点是其目标函数引入了正则化项^[25]。

3 结果与分析

3.1 土壤样品铬和锌含量分析

土壤样品中的 Cr 含量统计分析如图 4 所示。土壤中 Cr 的含量范围为 48~250 mg/kg, 平均含量为 150 mg/kg。只有 1% 的土壤样本处于 90 mg/kg 以下, 其余土壤样本 Cr 含量都超过 90 mg/kg (土

壤环境质量标准中维持自然背景的土壤质量的限制值)。大约有6%的土壤样本的Cr元素含量超过200 mg/kg。

土壤样品中的Zn含量统计分析如图5所示。土壤中Zn的含量范围为70~368 mg/kg, 平均含量为150 mg/kg。只有不到9%的土壤样本处于100 mg/kg以下, 其余土壤样本Zn含量都超过100 mg/kg(土壤环境质量标准中维持自然背景的土壤质量的限制值)。

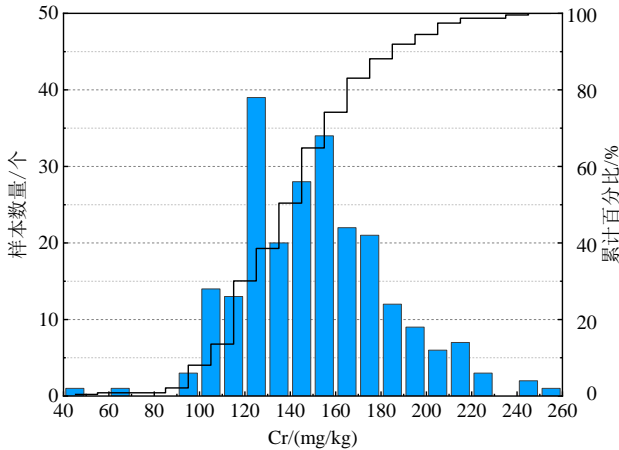


图4 土壤样品中Cr含量统计分析

Fig. 4 Statistical analysis of Cr content in soil samples

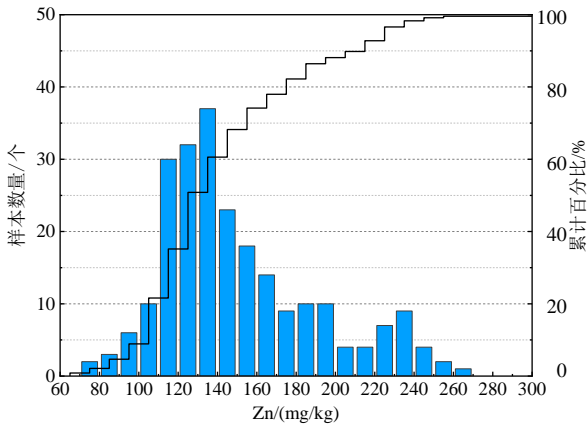


图5 土壤样品中Zn含量统计分析

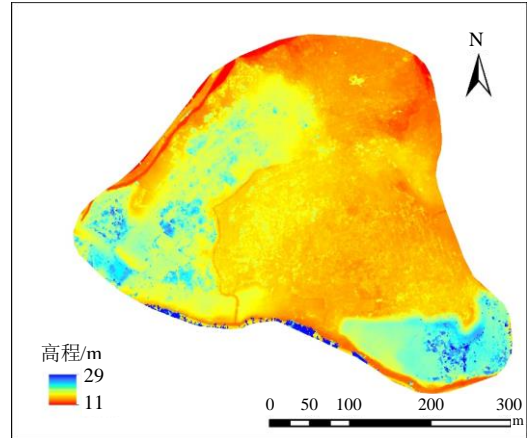
Fig. 5 Statistical analysis of Zn content in soil samples

3.2 场地地形因子提取与相关性分析

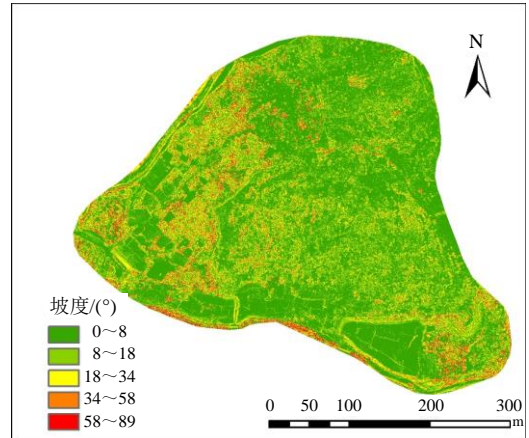
通过无人机拍摄获得的该场地的地形数据影像如图6所示。图6(a)为场地的高程分布图, 该场地的东南位置与西南位置呈现较高的地势。该场地的坡度分布图如图6(b)所示, 坡向分布图如图6(c)所示。为了综合反映地形特征, 本研究引用地形位指数, 其计算方法如下:

$$T = \log_{10} \left[\left(\frac{E}{E_0} + 1 \right) \times \left(\frac{S}{S_0} + 1 \right) \right] \quad (1)$$

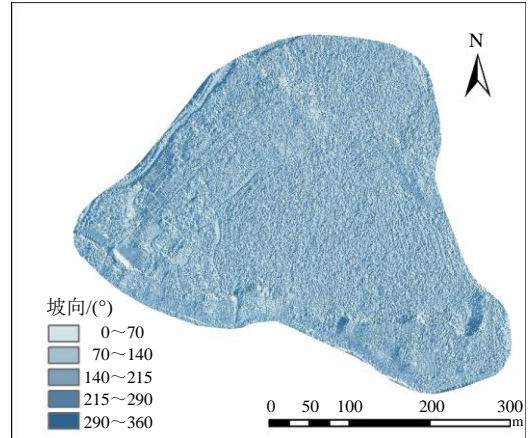
式中: T 为地形位指数; E 为采样点高程; E_0 为研究区平均高程; S 为采样点坡度; S_0 为研究区平均坡度。



(a) 高程



(b) 坡度



(c) 坡向

图6 场地的地形数据影像

Fig. 6 Terrain images of the site

该场地的地形位指数在0.01~1.06之间。在本研究中将地形位指数分为10个等级, 1级: $0 < T \leq 0.1$; 2级: $0.1 < T \leq 0.2$; 3级: $0.2 < T \leq 0.3$; 4级: $0.3 < T \leq 0.4$; 5级: $0.4 < T \leq 0.5$; 6级: $0.5 < T \leq 0.6$; 7级: $0.6 < T \leq 0.7$; 8级: $0.7 < T \leq 0.8$; 9级: $0.8 < T \leq 0.9$; 10级: $0.9 < T$ 。

土壤中 Cr 含量与地形位指数等级相关性如图 7 所示。土壤中 Zn 含量与地形位指数等级相关性如图 8 所示。土壤中 Cr 含量在 1 级地形位指数的平均值为 178.5 mg/kg, 而在 10 级地形位指数的平均值为 117.6 mg/kg。土壤中 Zn 含量在 1 级地形位指数的平均值为 213.5 mg/kg, 而在 10 级地形位指数的平均值为 112.6 mg/kg。

土壤 Cr 含量在地形位指数 1~3 级和 8~10 级间变化较大, 在 3~8 级间变化较平缓。土壤 Zn 含量在地形位指数 1~2 级间变化较大, 在其余级间变化较平缓。总体上, Cr 和 Zn 这两种重金属元素在土壤中的含量随着地形位指数等级的增高而降低。说明在场地地势较低洼的地方, 可能由于地表径流的携带作用更加容易出现重金属的聚集, 同时也说明了地形因子与场地土壤中重金属的分布存在一定的相关性。

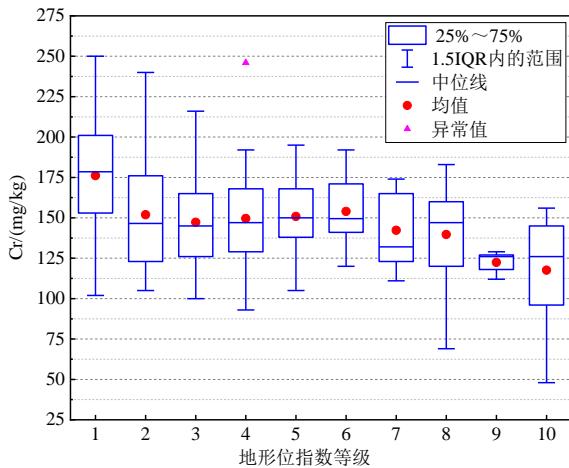


图 7 土壤中 Cr 含量与地形位指数等级相关性

Fig. 7 Correlation between the soil Cr content and grade of terrain niche

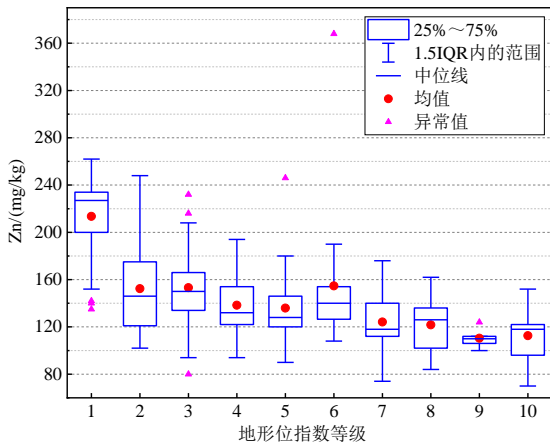


图 8 土壤中 Zn 含量与地形位指数等级相关性

Fig. 8 Correlation between the soil Zn content and grade of terrain niche

3.3 土壤重金属预测效果评估

为了评价模型的预测效果, 选取了两个评估指标: 均方根误差 (Root Mean Square Error, RMSE) 和平均绝对误差 (Mean Absolute Error, MAE), 定义为:

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (observed_i - predicted_i)^2}{N}} \quad (2)$$

$$MAE = \frac{\sum_{i=1}^N |observed_i - predicted_i|}{N} \quad (3)$$

式中: *observed* 是真实值; *predicted* 是预测值; *N* 是样本数量。

经过训练后, 建立土壤中 Cr 和 Zn 含量的预测模型。根据 RMSE 和 MAE 对 XGBoost、RF、和 GBDT 模型的性能效果进行了测试。RMSE 和 MAE 越小, 模型表现越好^[26]。同时本文将这些预测模型的效果与传统的克里金插值法进行了对比。

不同算法预测土壤 Cr 含量误差分析如图 9 所示。总体而言, GBDT 对土壤中 Cr 浓度的预测效果最好, 对应的 MAE 为 12.4 mg/kg, 对应的 RMSE 为 17.6 mg/kg。预测效果其次是 RF, 对应的 MAE 为 13.2 mg/kg, 对应的 RMSE 为 18.1 mg/kg。3 种机器学习预测模型最差的是 XGBoost, 对应的 MAE 为 15.3 mg/kg, 对应的 RMSE 为 19.5 mg/kg。但是 3 种机器学习预测模型对于 Cr 浓度的预测效果都比克里金插值法好。GBDT 作为预测效果最好的预测模型, 相对于克里金插值法, MAE 降低了 33.6%, RMSE 降低了 29.6%。

不同算法预测土壤 Zn 含量误差分析如图 10 所示。与土壤中 Cr 含量的预测效果类似, GBDT 对土壤中 Zn 含量的预测效果最好, 对应的 MAE 为 13.6 mg/kg, 对应的 RMSE 为 18.8 mg/kg。预测效果其次是 RF, 对应的 MAE 为 16.1 mg/kg, 对应的 RMSE 为 21.5 mg/kg。3 种机器学习预测模型最差的是 XGBoost, 对应的 MAE 为 19.2 mg/kg, 对应的 RMSE 为 24.5 mg/kg。与土壤中 Cr 含量的预测效果类似, 3 种机器学习预测模型对于 Zn 含量的预测效果都比克里金插值法好。GBDT 作为预测效果最好的预测模型, 相对于克里金插值法, MAE 降低了 52.1%, RMSE 降低了 51.2%。

XGBoost 作为一种提升树模型, 用一棵树去预测一个值, 得到该值与实际值的偏差, 再添加一

棵树去学习该偏差^[27]。RF是由独立的决策树组成的一个森林，在最终模型组合时，每棵树的权重相等^[28]。GBDT回归算法是一种集成算法，通过对决策树的合并来构建一个更强大的模型，该算法采用连续的方式进行树的构造，后一棵树纠正前一棵树的错误，每次迭代使损失函数下降^[29]。GBDT回归算法背后的主要思想是许多简单模型的合并，表现出强大的预测性能。同样在本研究中GBDT回归算法有更好的预测效果。另一方面，GBDT与XGBoost和RF的特征选取策略不同，GBDT遍历所有特征，XGBoost和RF则进行特征的子采样来加快运算速度。3种算法特征选取策略的不同也是导致算法预测效果不同的一个重要因素。以评估效果最好的GBDT为例，进行具体点位的误差分析。对于土壤Cr含量的预测，GBDT算法对土壤Cr含量预测误差分布如图11所示。GBDT对土壤Cr含量的预测误差范围为0.2~53.6 mg/kg。预测误差最大的点是235号取样点；

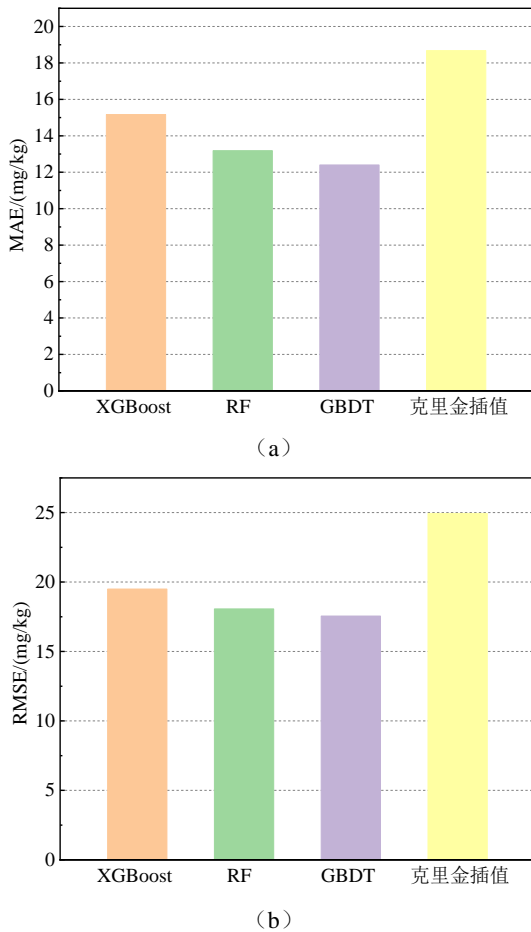


图9 不同算法预测土壤Cr含量误差分析
Fig. 9 Error analysis of different algorithms for predicting the soil Cr content

预测误差最小的点是47号取样点，预测结果与真实结果几乎一致。只有18%的点位预测的误差大于20 mg/kg，其余点位的误差全部小于20 mg/kg，58%的土壤点位的误差小于10 mg/kg。

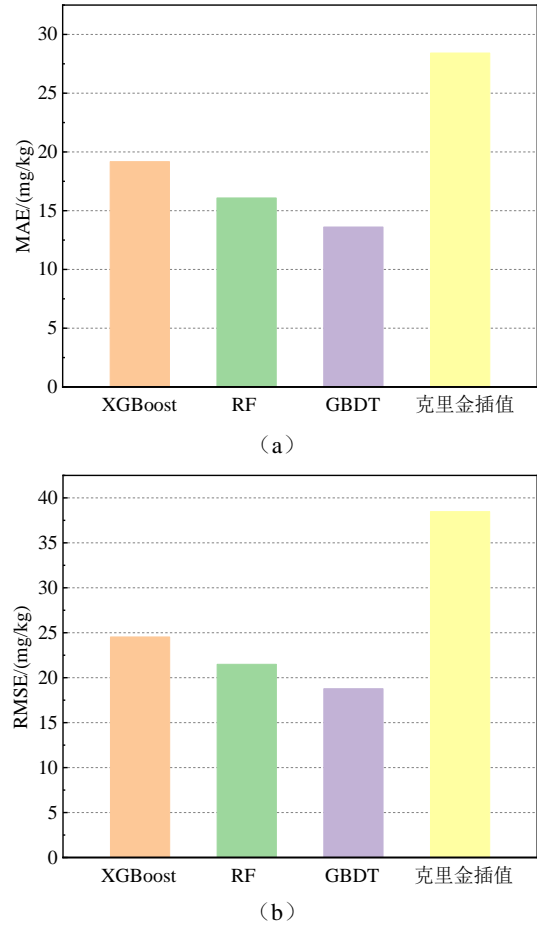


图10 不同算法预测土壤Zn含量误差分析
Fig. 10 Error analysis of different algorithms for predicting the soil Zn content

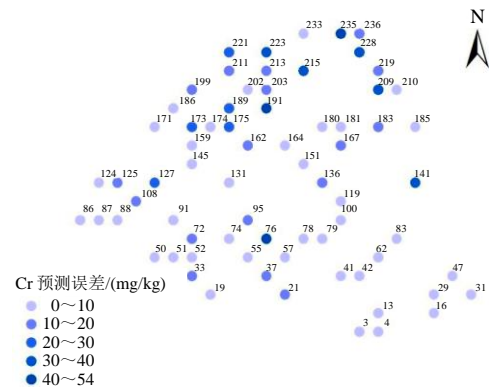


图11 GBDT算法对土壤Cr含量预测误差分布
Fig. 11 Prediction error distribution of soil Cr content predicted by GBDT algorithm

对于土壤Zn含量的预测，GBDT算法对土壤Zn含量预测误差分布如图12所示。GBDT对土壤Zn含量的预测误差范围为0.001~58.600 mg/kg。

预测误差最大的点是 171 号取样点, 预测误差最小的点是 41 号取样点, 预测结果与真实结果几乎一致。大致有 22.5% 的点位预测的误差大于 20 mg/kg, 其余点位的误差全部小于 20 mg/kg。与 Cr 类似, 56% 的土壤点位的误差小于 10 mg/kg。

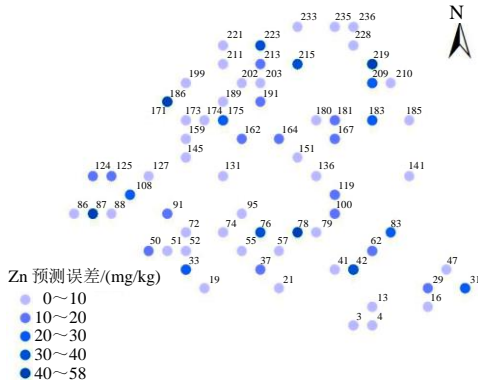


图 12 GBDT 算法对土壤 Zn 含量预测误差分布
Fig. 12 Prediction error distribution of soil Zn content predicted by GBDT algorithm

3.4 场地重金属预测特征因子贡献度分析

XGBoost、RF 和 GBDT 可以量化每个特征变量在预测中的相对重要性, 并用于确定影响重金属含量因素的贡献程度^[21,23]。本文以评估效果最好的 GBDT 模型为例, 进行特征因子贡献度的分析。

图 13 显示 GBDT 模型中, 各个特征因子对于土壤 Cr 含量预测的贡献度。

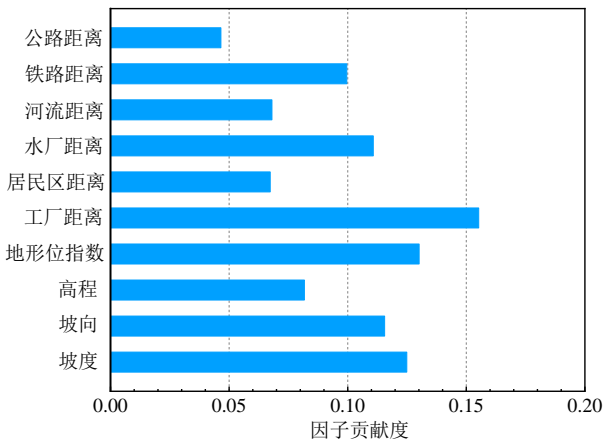


图 13 土壤 Cr 含量预测因子贡献度分析
Fig. 13 Factor contribution analysis of soil Cr content prediction

10 个特征因子中工厂距离对于 Cr 含量的预测贡献度最大, 贡献度为 15.5%, 因此可以推测研究区土壤中 Cr 的污染很大可能受到工厂的影响。地形位指数和坡度的贡献度分别位于第二位和第三位, 对应的贡献度分别为 13.0% 和 12.5%。因此可

以推测场地土壤 Cr 含量的分布同时受到地形地貌的影响, 比如地表径流受到地形地貌的影响, 雨水等携带重金属在此处聚集, 进而影响地表土壤重金属的迁移与分布。

土壤 Zn 含量预测因子贡献度分析如图 14 所示。在潜在污染源距离因子中, 工厂与采样点的距离对于 Zn 含量的预测贡献度最大, 为 15.1%, 因此推测土壤中 Zn 富集很大的原因来自工厂。在地形类因子中, 坡向占有最大的贡献度, 对应的贡献度为 22.3%。

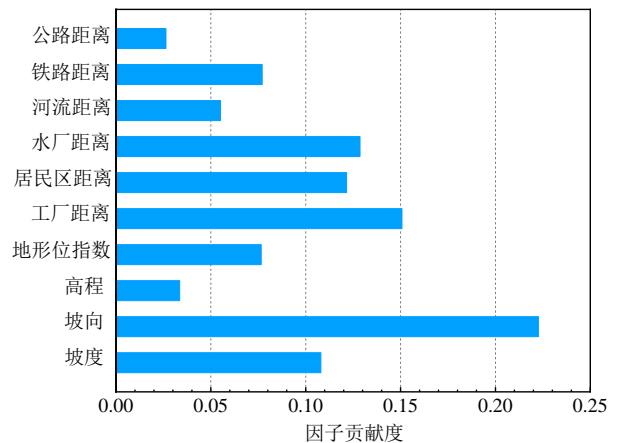


图 14 土壤 Zn 含量预测因子贡献度分析
Fig. 14 Factor contribution analysis of soil Zn content prediction

4 结 论

本文以安徽某重金属污染场地的土壤为研究对象, 采集土壤样本 236 个, 并获得土壤中重金属 Cr 和 Zn 的含量。对场地使用无人机进行航测获得场地影像及地形模型, 提取无人机影像中的地形因子及采样点与周边潜在污染源的因子作为预测模型自变量, 采用 GBDT、RF、XGBoost 这 3 种机器学习回归算法进行土壤 Cr 和 Zn 含量的预测。本文主要的结论如下:

(1) 场地土壤 Cr 和 Zn 含量随着地形位指数等级的增高而降低, 说明在场地地势较低洼的地方更加容易出现重金属的富集, 同时也说明了地形因子与场地中重金属的分布存在一定的相关性。

(2) 3 种机器学习回归算法对土壤重金属的预测有很好的效果。其中 GBDT 对土壤中 Cr 和 Zn 含量的预测效果最好, 对土壤中 Cr 含量预测对应的 MAE 为 12.4 mg/kg, 对应的 RMSE 为 17.6 mg/kg; 对土壤中 Zn 含量预测对应的 MAE 为 13.6 mg/kg, 对应的 RMSE 为 18.8 mg/kg。

(3) 3种机器学习预测模型对于 Cr 和 Zn 含量的预测效果都比克里金插值好。其中 GBDT 相对于克里金插值法, 对于 Cr 含量的预测 MAE 降低了 33.6%, RMSE 降低了 29.6%; 对于 Zn 含量的预测 MAE 降低了 52.1%, RMSE 降低了 51.2%。

(4) 在选取的 10 个特征因子中, 工厂距离对于 Cr 和 Zn 含量的预测贡献度较大, 贡献度分别为 15.5% 和 15.1%, 因此可以推测研究区土壤 Cr 和 Zn 的污染很大可能来源于工厂。

本研究存在一些局限性, 研究使用的地形模型未能排除地表植物覆盖的影响, 在未来研究中可以考虑使用机载 LiDAR 进行场地地形模型的构建, 进而消除植被覆盖对于地形数据的影响。许多工业污染场地更关注深度方向的污染问题, 而本研究采集的样本全部为地表土壤样本, 模型的训练基于地表土壤样本数据, 且无人机获取的影像特征为地表土壤污染的相关影响特征因子, 不足以支撑深度方向上土壤污染的预测。因此本研究方法对深度方向土壤污染的预测适用性不足。

参考文献

- [1] OVES M, KHAN M S, ZAIDI A, et al. Soil contamination, nutritive value, and human health risk assessment of heavy metals: an overview[M]//ZAIDI A, WANI P A, KHAN M S. Toxicity of Heavy Metals to Legumes and Bioremediation. Vienna: Springer, 2012: 1-27.
- [2] FACCHINELLI A, SACCHI E, MALLEN L. Multivariate statistical and GIS-based approach to identify heavy metal sources in soils[J]. Environmental Pollution, 2001, 114(3): 313-324.
- [3] MURTAZA G, USMAN Y, NIAZI N K, et al. Bioaccumulation of potentially toxic elements in cereal and legume crops: a review[J]. Clean-Soil Air Water, 2017, 45(12): 1700548.
- [4] SOLGI E, ESMALI-SARI A, RIYAH-BAKHTIARI A, et al. Soil contamination of metals in the three industrial estates, Arak, Iran[J]. Bulletin of Environmental Contamination and Toxicology, 2012, 88(4): 634-638.
- [5] NOLI F, TSAMOS P. Concentration of heavy metals and trace elements in soils, waters and vegetables and assessment of health risk in the vicinity of a lignite-fired power plant[J]. Science of the Total Environment, 2016, 563: 377-385.
- [6] YADAV I C, DEVI N L, SINGH V K, et al. Spatial distribution, source analysis, and health risk assessment of heavy metals contamination in house dust and surface soil from four major cities of Nepal[J]. Chemosphere, 2019, 218: 1100-1113.
- [7] ZHAO K L, LIU X M, XU J M, et al. Heavy metal contaminations in a soil-rice system: Identification of spatial dependence in relation to soil properties of paddy fields[J]. Journal of Hazardous Materials, 2010, 181(1-3): 778-787.
- [8] CAO S Z, DUAN X L, ZHAO X G, et al. Health risks from the exposure of children to As, Se, Pb and other heavy metals near the largest coking plant in China[J]. Science of the Total Environment, 2014, 472: 1001-1009.
- [9] GALLINI L, AJMONE-MARSAN F, SCALENGHE R. The contamination legacy of a decommissioned iron smelter in the Italian Alps[J]. Journal of Geochemical Exploration, 2018, 186: 121-128.
- [10] CAPOLUPO A, PINDOZZI S, OKELLO C, et al. Photogrammetry for environmental monitoring: the use of drones and hydrological models for detection of soil contaminated by copper[J]. Science of the Total Environment, 2015, 514: 298-306.
- [11] HOU D Y, O'CONNOR D, NATHANAIL P, et al. Integrated GIS and multivariate statistical analysis for regional scale assessment of heavy metal soil contamination: A critical review[J]. Environmental Pollution, 2017, 231: 1188-1200.
- [12] DUBIN R A. Spatial autocorrelation and neighborhood quality[J]. Regional Science and Urban Economics, 1992, 22(3): 433-452.
- [13] WU Y Z, ZHANG X, LIAO Q L, et al. Can contaminant elements in soils be assessed by remote sensing technology: a case study with simulated data[J]. Soil Science, 2011, 176(4): 196-205.
- [14] LIU M L, WANG T J, SKIDMORE A K, et al. Heavy metal-induced stress in rice crops detected using multi-temporal Sentinel-2 satellite images[J]. Science of the Total Environment, 2018, 637: 18-29.
- [15] TAN K, WANG H M, CHEN L H, et al. Estimation of the spatial distribution of heavy metal in agricultural soils using airborne hyperspectral imaging and random forest[J]. Journal of Hazardous Materials, 2020, 382: 120987.

- [16] SHI T Z, CHEN Y Y, LIU Y L, et al. Visible and near-infrared reflectance spectroscopy—an alternative for monitoring soil contamination by heavy metals[J]. *Journal of Hazardous Materials*, 2014, 265: 166–176.
- [17] 李国旭, 耿静, 许选虹, 等. 基于 WorldView-3 多光谱和关键环境变量的土壤镉含量反演[J]. *农业工程学报*, 2022, 38(12): 224–232.
- LI Guo-xu, GENG Jing, XU Xuan-hong, et al. Inversion of soil Cd content using WorldView-3 multispectral and key environmental variables[J]. *Transactions of the Chinese Society of Agricultural Engineering*, 2022, 38(12): 224–232.
- [18] 尹芳, 封凯, 吴朦朦, 等. 一种基于分段偏最小二乘模型的土壤重金属遥感反演方法[J]. *遥感技术与应用*, 2021, 36(6): 1321–1328.
- YIN Fang, FENG Kai, WU Meng-meng, et al. A remote sensing estimation method for heavy metals in soil based on piecewise partial least squares model[J]. *Remote Sensing Technology and Application*, 2021, 36(6): 1321–1328.
- [19] 王腾军, 方珂, 杨耘, 等. 随机森林回归模型用于土壤重金属含量多光谱遥感反演[J]. *测绘通报*, 2021(11): 92–95.
- WANG Teng-jun, FANG Ke, YANG Yun, et al. Multi-spectral remote sensing inversion of soil heavy metal content using random forest regression model[J]. *Bulletin of Surveying and Mapping*, 2021(11): 92–95.
- [20] VÅGEN T G, WINOWIECKI L A, TONDOH J E, et al. Mapping of soil properties and land degradation risk in Africa using MODIS reflectance[J]. *Geoderma*, 2016, 263: 216–225.
- [21] LIANG W Z, LUO S Z, ZHAO G Y, et al. Predicting hard rock pillar stability using GBDT, XGBoost, and LightGBM algorithms[J]. *Mathematics*, 2020, 8(5): 765.
- [22] MOISEN G G, FREEMAN E A, BLACKARD J A, et al. Predicting tree species presence and basal area in Utah: a comparison of stochastic gradient boosting, generalized additive models, and tree-based methods[J]. *Ecological Modelling*, 2006, 199(2): 176–187.
- [23] BREIMAN L. Random forests[J]. *Machine Learning*, 2001, 45(1): 5–32.
- [24] QIU Y, ZHOU J, KHANDELWAL M, et al. Performance evaluation of hybrid WOA-XGBoost, GWO-XGBoost and BO-XGBoost models to predict blast-induced ground vibration[J]. *Engineering with Computers*, 2022, 38(S5): 4145–4162.
- [25] TAN K, MA W B, CHEN L H, et al. Estimating the distribution trend of soil heavy metals in mining area from HyMap airborne hyperspectral imagery based on ensemble learning[J]. *Journal of Hazardous Materials*, 2021, 401: 123288.
- [26] ZHENG G H, RYU D, JIAO C X, et al. Estimation of organic matter content in coastal soil using reflectance spectroscopy[J]. *Pedosphere*, 2016, 26(1): 130–136.
- [27] 李静波, 张莹, 盖荣丽. 基于机器学习的星载短波红外 CO₂ 柱浓度估算研究[J]. *中国环境科学*, 2023, 43(4): 1499–1509.
- LI Jing-bo, ZHANG Ying, GAI Rong-li. Estimation of the column concentration of carbon dioxide using spaceborne shortwave infrared spectrometer[J]. *China Environmental Science*, 2023, 43(4): 1499–1509.
- [28] 王可心, 包云轩, 朱承瑛, 等. 随机森林回归法在冬季路面温度预报中的应用[J]. *气象*, 2021, 47(1): 82–93.
- WANG Ke-xin, BAO Yun-xuan, ZHU Cheng-ying, et al. Forecasts of road surface temperature in winter based on random forests regression[J]. *Meteorological Monthly*, 2021, 47(1): 82–93.
- [29] 韩启迪, 张小桐, 申维. 基于梯度提升决策树(GBDT)算法的岩性识别技术[J]. *矿物岩石地球化学通报*, 2018, 37(6): 1173–1180.
- HAN Qi-di, ZHANG Xiao-tong, SHEN Wei. Lithology identification technology based on gradient boosting decision tree (GBDT) algorithm[J]. *Bulletin of Mineralogy, Petrology and Geochemistry*, 2018, 37(6): 1173–1180.