

遗传算法优化神经网络的重金属污染源预测

辛艾莹

中南安全环境技术研究院股份有限公司 湖北 武汉 430000

摘要: 城市土壤重金属污染日益严重,而确定重金属污染源位置是一个较为突出的难题。本文以武汉市为案例研究区,利用2017年获取的100个土壤重金属浓度采样数据,选取Zn、Cu、Mn、B四种重金属指标,分析武汉市重金属空间分布情况,并构建一种基于遗传算法优化的BP神经网络(GA-BPNN)模型,用于预测土壤重金属污染源位置。得到结论如下:(1)武汉市重金属浓度呈中心点浓度最高向外扩散渐渐降低的点源式分布,重金属污染为点源污染,且Zn、B污染主要集中在中心城区,Cu、Mn污染主要集中在偏远城区;(2)GA-BPNN模型表现出良好的训练效果和较高的预测精度,检验组数据的拟合精度超过80%。对于硼(B)元素的污染源预测,模型预测位置与实测最大浓度点位置完全一致,浓度预测误差在7%以内。

关键词: 重金属;遗传算法;BP神经网络;污染源寻源预测

1 引言

土壤中的重金属污染具有持久性和隐蔽性。这些污染物一旦进入土壤,由于其不能被微生物有效分解,会在土壤中不断累积,改变土壤性质^[1]。某些重金属还可能在特定条件下转化为毒性更强的有机化合物,如烷基化合物。这些污染物易被植物及其他生物体吸收和富集,并通过食物链逐级传递,最终在人体或牲畜体内蓄积,对植物生长、动物健康乃至人类生命安全构成直接或间接的威胁^[2]。重金属污染物并非仅停留在土壤中,它们还可以通过地表径流和渗透作用进入地表水和地下水系统,进一步扩大污染范围,造成水环境的重金属污染^[3]。本研究选取武汉市,分析Zn、Cu、Mn、B四种重金属分布,构建遗传算法优化神经网络模型,旨在实现污染源定位,为土壤污染防治修复提供科学依据,提升环境质量。

2 材料与方法

2.1 研究区域概况

本文选取武汉市作为案例研究区。研究区域辖七个中心城区(江岸、江汉、硚口、汉阳、武昌、青山、洪山)、六个远城区(东西湖、汉南、蔡甸、江夏、新洲和黄陂)及三个功能区(经济技术开发区、东湖新技术开发区、长江新区)。

2.2 数据获取与预处理

本文使用的2017年武汉市100个土壤表层采样点Zn、Cu、Mn、B四种重金属元素浓度实测值来源于湖北大学生态所资料室。重金属浓度值采用最小-最大规范法进行归一化处理,以消除不同重金属浓度量纲和数量级差异对模型训练的影响。其计算公式如下:

$$X_n = \frac{X_i - X_{min}}{X_{max} - X_{min}}$$

将100组样本数据按照4:1的比例随机划分为训练集(80组)和检验集(20组)。训练集用于GA-BPNN模型的构建和参数优化,检验集用于评估模型的泛化能力和检验精度。

2.3 空间分析-克里金插值

为了初步了解研究区内四种重金属的空间分布特征并确定主要污染源的搜索范围,本研究利用ArcGIS10.2软件将武汉市划分为1km*1km的区域,并采用克里金(Kriging)插值对100个采样点的重金属浓度数据进行空间分析。克里金法是一种基于空间自相关的地统计学最优内插方法,能够提供无偏最优估计,并量化预测不确定性,广泛应用于土壤属性制图和污染物空间分布研究^[4]。

2.4 GA-BPNN模型构建

本研究采用三层BP神经网络结构。输入层包含2个神经元,分别对应采样点的网格坐标(X,Y);隐含层包含1层,神经元数量根据经验公式 $n1 = \sqrt{n + m} + a$ 确定,其中n为输入层神经元数(2),m为输出层神经元数(4),a取常数5,故隐含层神经元数为8;输出层包含4个神经元,分别对应四种重金属的预测浓度值。网络中神经元激活函数选取tanh函数,学习算法采用Adam优化器,训练目标是最小化预测输出与实际观测值之间的误差,误差函数定义为均方误差。

遗传算法的编码采用二进制编码方式对BP神经网络的权值和阈值进行编码形成个体,种群规模设定为100,最大进化代数设定为100代,交叉概率设定为0.3,变异

概率设定为0.2, 适应度函数定义为BP神经网络预测误差E的倒数。通过计算每个个体的适应度, 采取选择、交叉、变异等遗传操作不断进化种群, 最终得到最优权值和阈值组合用于BP神经网络后续的训练及预测^[5]。

3 结论

3.1 武汉市土壤重金属空间分布特征

Zn、Cu、Mn、B四种重金属浓度空间分布见图1。武汉市土壤重金属含量总体较高, 且大部分重金属的浓度分布呈现多个点源式特征, 即以若干中心点浓度最高, 向外逐渐扩散降低。这初步表明武汉市土壤重金属污染可能主要来源于集中的点源排放, 如特定工厂或商业活动区域。

结合城市行政区划与人口密度分析发现, 人口密度

较大的区域, 其土壤重金属浓度通常高于人口密度较小的区域, 暗示交通、商业、工业等强烈的人为活动是武汉市土壤重金属的主要来源。区域对比显示, 郊区的Cu、Mn浓度高于中心城区, 这可能与部分工厂搬迁至郊区有关; 而Zn作为交通污染的标志性元素, 其浓度在中心城区明显高于其他区域。在中心城区内部, 武昌区和洪山区由于分别是省政府所在地、文化教育中心以及城市化进程较晚的高校集中地, 其总体土壤重金属水平相对较低。相反, 人口密集、商业繁华的江汉区以及作为老工业基地的青山区、汉阳区, 其土壤重金属含量则总体较高。这些初步的空间分析为后续GA-BPNN模型的污染源搜索范围划定提供了重要参考。

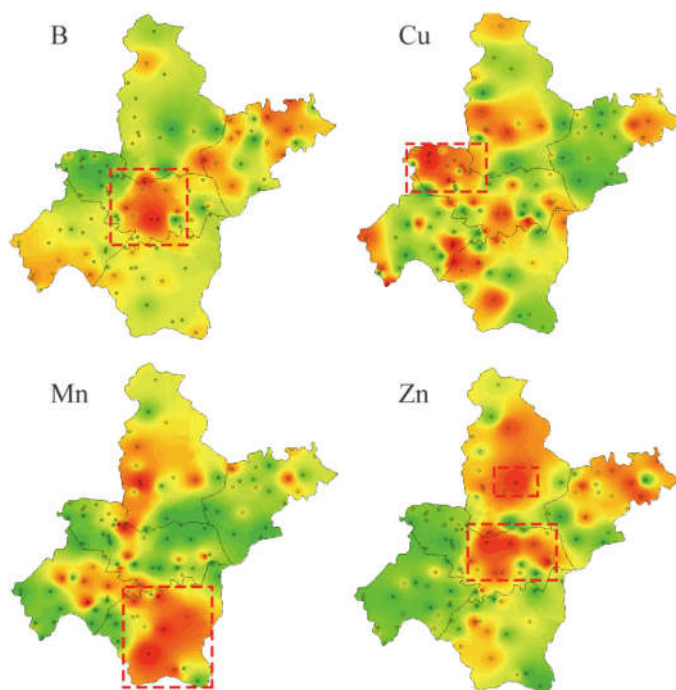


图1 武汉市土壤重金属浓度空间分布图

3.2 GA-BPNN模型性能评估

GA-BPNN模型经过1000次迭代训练后, 其误差变化如图2.a所示。训练过程中, 所有训练数据的平均误差、最大误差和最小误差均呈现向误差减小趋势演进。与初始迭代相比, 最终迭代时数据的平均误差和最小误差均低于20%, 最大误差也接近20%, 三种误差均减小了40%以上。

最终代训练结果的误差统计分析结果显示(见图2.b), 对于每种重金属(Zn、Cu、Mn、B), 均有超过60%的训练样本预测误差在20%以内。除极少数样本(1.25%)误差超过40%外, 其余所有样本的预测误差均

控制在40%以内。每种重金属预测值的准确率均在80%以上。从所有重金属数据整体来看, 61.825%的训练数据预测误差在20%以内。这些结果表明, GA-BPNN模型在训练阶段达到了良好的收敛效果, 误差被有效控制在合理范围内。

将未经训练的20组检验数据输入到训练好的GA-BPNN模型中, 得到的预测误差分布如图2.c所示。检验数据显示, 大部分预测值与真实值的误差在20%以内, 仅有3个数据点的误差超过40%, 整体误差分布与训练数据相似。这表明该模型不仅对训练样本有效, 而且具有良好的泛化能力, 能够对新数据进行较为准确的预测。

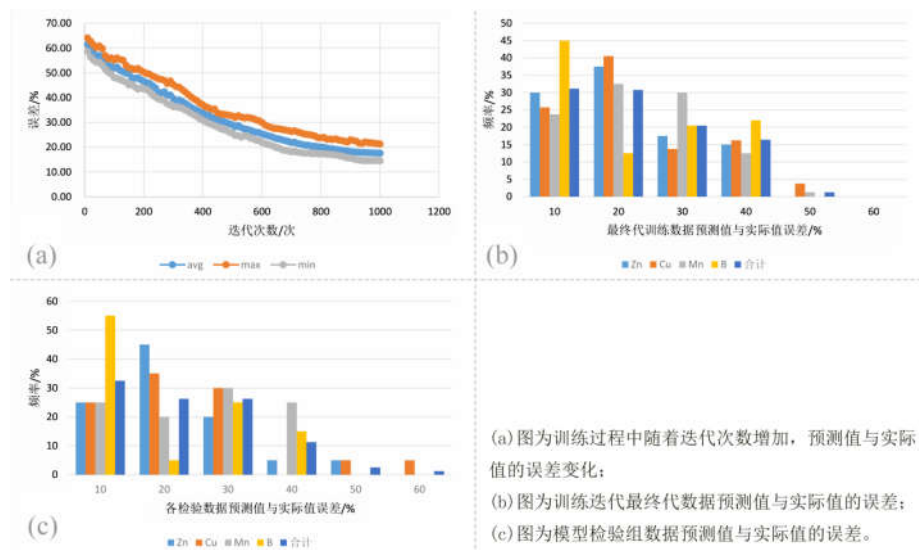


图2 模型误差分析图

3.3 武汉市土壤重金属主要污染源位置预测

根据图3.1确定主要污染源位置范围是在图中红色区域内。因此，可将模型的搜索范围定为图中红色较重区域。对于Zn，搜索范围为[(45,54)，(86,137)]、[(47,87)，(80,113)]；对于Cu，搜索范围为[(21,67)，(53,93)]；对于Mn，搜索范围为[(48,3)，(84,47)]；对于B，搜索范围为[(46,46),(74,81)]。

根据上诉搜索范围，采用遗传算法优化神经网络模型，得到了可能的污染源位置及对应的浓度值。具体数值如表1所示。污染源预测位置位于图中红色最深部分，与采样点最大值所对应的的坐标相近，浓度接近或超过采样点浓度最大值。同时，B元素重金属污染源预测位置与采样点B元素浓度最大值所对应位置相同，并且预测的浓度误差在7%以内，证明该预测结果具有一定可靠性。

表1 重金属主要污染源预测位置

重金属类型	坐标	浓度值
Zn	(55,71)	0.956
	(67,101)	1.038
Cu	(29,84)	1.108
Mn	(91,9)	1.192
B	(57,55)	0.934

4 讨论

(1) GA-BPNN模型表现出良好的收敛性和较高的预测精度。通过遗传算法优化BP神经网络的初始权值和阈值，有效提升了模型的全局搜索能力，避免了传统BP网络易陷入局部最优的问题。

(2) 该模型能够较为准确地定位研究区内主要重金属污染源。B的预测准确率相对于其他三种金属准确率更

高，这可能因为B属于类重金属，其本身特性较其他三种重金属而言不同，对预测结果有影响；Cu和Mn的预测准确率则相对较低,可能是因为其污染源分布更为复杂，拥有多个点源，多个源之间会相互影响导致预测精度相对较低。

(3) 重金属的污染可能具有聚集性，各个重金属之间会有相互影响，从而使得一种重金属的污染源可能和另一种重金属污染源具有相关性，但是该模型并没有考虑各种重金属污染源之间的关联性，因此，之后如果需要对模型进行改进可以将重金属之间关联性纳入考虑范围。

(4) 尽管存在局限，GA-BPNN模型作为一种智能化的污染源初步识别工具，在数据相对稀疏的情况下仍能提供有价值的线索，可为环境管理部门进行后续的详细调查、精准监测和污染治理提供科学依据和技术支持。

参考文献

[1]岳瑞,张梦林,段佳辉,等.重庆市典型生态功能区土壤重金属污染评价与源解析[J].安全与环境学报,2025,25

[2]王亚婷,廖明旭,康馨文,等.某电镀企业土壤重金属污染特征及健康风险评价[J].安全与环境学报,2024,24(11):4464-4474.

[3]徐剑乔.基于改进主成分分析法的水环境中重金属污染流域水质生态评价方法研究[J].环境科学与管理,2024,49(5):191-194.

[4]霍霄妮,李红,张微微,等.北京耕作土壤重金属多尺度空间结构[J].农业工程学报,2009,25(3):223-229.

[5]杨娟,王昌全,李冰,等.基于BP神经网络的城市边缘带土壤重金属污染预测—以成都平原土壤Cd为例[J].土壤学报,2008,44(3):430-436.