基于 GA-SVM 的耕地土壤重金属含量 高光谱反演方法的研究

郭云开1,2,张思爱1,2*,谢晓峰1,2,谢 琼3,4

(1.长沙理工大学交通运输工程学院 湖南 长沙 410076; 2.长沙理工大学测绘遥感应用技术研究所,湖南 长沙 410076; 3.湖南工程职业 技术学院测绘地理学院,湖南 长沙 410151; 4.长沙理工大学公路地质灾变预警空间信息技术湖南省工程实验室,湖南 长沙 410114)

摘 要:为提高耕地土壤重金属含量高光谱反演模型精度,以岳阳县某地区耕地土壤重金属铁(Fe)、砷(As)、铬(Cr)为例,提出了一种遗传算法(GA)优化支持向量机(SVM)的重金属含量反演模型。在对光谱进行 SG 平滑和 10 nm 重采样后,利用一阶/二阶微分、倒数对数和连续统去除光谱变换方法增强光谱特征,通过相关性分析筛选最优变换光谱,使用皮尔森相关系数与主成分分析提取各重金属光谱特征变量,分别建立 SVM 和 GA-SVM 土壤重金属高光谱 反演模型并进行精度验证。结果表明,二阶微分变换光谱与各重金属含量相关性整体最突出;三种重金属在可见光波段 490 nm、500 nm、510 nm 和 530 nm 具有共同敏感特征;经 GA 算法优化 SVM 参数后,对比 SVM 回归模型,预测精度 有明显提高,其重金属 Fe、As 和 Cr 的验证集 *R*²分别为 0.968、0.821 和 0.976;研究结果可为应用遥感技术反演耕地土 壤重金属含量提供新的参考。

关 键 词: 土壤重金属; 高光谱遥感; 光谱变换; 遗传算法; 支持向量机 中图分类号: TP79 **文献标识码:** A **文章编号:** 0564-3945(2021)04-0968-07

DOI: 10.19336/j.cnki.trtb.2020062201

郭云开, 张思爱, 谢晓峰, 谢 琼. 基于 GA-SVM 的耕地土壤重金属含量高光谱反演方法的研究 [J]. 土壤通报, 2021, 52(4): 968 – 974

GUO Yun-kai, ZHANG Si-ai, XIE Xiao-feng, XIE Qiong. The Hyperspectral Inversion Method of Heavy Metal Contents in Cultivated Soils Based on GA-SVM[J]. Chinese Journal of Soil Science, 2021, 52(4): 968 – 974

随着工业化和城镇化的快速发展,土壤重金属 污染问题变得日益严峻[1-2]。过量的 As 和 Cr 在人体 内富集将严重影响身体健康^[3]。传统土壤重金属含量 监测方法具有费用高、耗时长等缺点,且不利于大 面积土壤的快速监测与评估[4-5]。高光谱遥感技术具 有光谱分辨率高、连续性强和波段多等特点,能够 快速准确的获取土壤信息,为定量预测分析土壤重 (FD)、倒数对数(LR)和连续统去除法(CR)对 土壤原始光谱进行处理,获取土壤成分的特征波段, 运用逐步回归模型分析,发现基于 FD 变换光谱的模 型精度最高四。涂宇龙等通过对原始光谱进行预处理, 分别采用主成分分析和相关系数分析提取特征变量, 运用逐步回归法建立土壤重金属 Cu 含量反演模型, 取得较好的 Cu 含量估算能力, R² 为 0.86^[8]。传统建 模反演方法在处理非线性和高维数据问题时会表现 出明显的不足,而支持向量机(Support Vector Machines, SVM)可以较好地针对这类问题进行分析^[9-11]。沈强等通过 SVM 对土壤有机质含量进行建模分析,得出基于二阶微分(SD)的 SVM 模型反演精度最高其决定系数 R²为 0.89^[12]。吕杰等使用 SVM 构建水稻 Cd 含量高光谱反演模型,发现经小 波变换后预测精度得到提升,其预测模型 R² 为 0.8674^[13]。目前,虽然对土壤重金属含量反演进行了 大量研究,但由于高光谱数据波段信息多,数据冗余等问题严重影响了模型反演精度,对光谱变量的 合理选取能够降低模型复杂度、提高模型预测精度^[14]。现有的变量提取方法主要是与线性模型相结合,而 与 SVM 模型结合的相关研究相对较少。因此,利用 不同变量提取算法与非线性回归模型相结合,为提高土壤重金属含量高光谱反演精度提供新方法。

本文以岳阳县某地区耕地土壤重金属为研究对

收稿日期: 2020-06-22; 修订日期: 2021-03-21

基金项目:国家自然科学基金项目(41471421,41671498)、长沙理工大学公路地质灾变预警空间信息技术湖南省工程实验室开放基金项目 (kfj190603)资助

作者简介:郭云开(1958-),湖南省常德市人,男,博士,教授,主要研究方向为高等级路域环境遥感与土壤环境遥感。E-mail: guoyunkai226@163.com

^{*}通讯作者: E-mail: 3393838418@qq.com

象,通过对原始反射光谱(R)数据先后进行 SG 平 滑和 10 nm 重采样,利用 FD、SD、LR 和 CR 光谱 变换方法增强光谱数据与土壤重金属的相关性,再 使用皮尔森相关系数和主成分分析对变换光谱(SD) 进行特征提取,以提取的特征变量作为光谱参量, 分别建立 SVM 和 GA-SVM 耕地土壤重金属含量高 光谱反演模型,对比不同特征提取下的两种反演模 型,寻找各重金属含量最优反演路径。

1 数据与方法

1.1 土壤采集与光谱测量

研究区位于湖南省岳阳县某地区,其地势平坦, 土壤肥沃,气候适宜,年均降水量为1200~1300 mm。 研究区总面积137.9 km²,具有丰富的农业资源,耕 地面积38.7 km²,以种植水稻为主,其农田面积 34.6 km²,是该区域的重要粮食产区。近年来,该地 区大力调整产业结构,在工业上大力招商引资,目 前已有工业企业78 家,在工业化快速发展的同时, 也产生了大量的重金属污染物与其他废弃物的排放, 导致该区域土壤环境受到严重破坏进而影响农产品

的生产。本此试验通过对研究区农田土壤进行均匀 取样,在土壤采集过程中,以样本光谱采集点为中 心,在其周围以"S"曲线确定5个土壤采样点。每点 采集土壤样品约 200 g,采样深度为 20 cm,装入带 有标签的深色塑料袋,本次实验共采集58个土壤样 本点。将土壤样本置于阳光处并保持通风使样本自 然风干,研磨样本并剔除其中杂质,然后使用100 目尼龙筛过滤,并储存在专用容器中。将最终获取 的每个样本分为两部分处理,一部分使用 AvaField-3地物波普仪获取土壤的光谱信息,其波段范围为 300~2500 nm, 在每个样本点中获取 10 条光谱曲线 并对其取平均来确定该点的原始光谱信息。剩余部 分通过专业检测机构测定土壤重金属 Fe、As 和 Cr 含量,各重金属含量的描述性统计如表1所示,其 中 Fe 元素含量变异系数为 0.210, As 元素含量变异 系数为 0.180, Cr 元素含量变异系数为 0.284。通过 土壤科学研究中变异系数对土壤性质变异程度的划 分标准:变异系数值在 0~0.1 为轻度变异,0.1~1 为中度变异,大于1为重度变异[14]。由此可知,研究 区土壤重金属 Fe、As 和 Cr 元素均属于中度变异。

表1 土壤重金属含量描述性统计 Table 1 Descriptive statistics of soil beauty metal content

| Table 1 Descriptive statistics of son neavy metal contents | | | | | | | | |
|--|---------|---------|---------|--------------------|--------------------------|--|--|--|
| 重金属 | 最小值 | 最大值 | 均值 | 标准差 | 变异系数 | | | |
| Heavy metal | Minimum | Maximum | Average | Standard deviation | Coefficient of variation | | | |
| Fe (g kg ⁻¹) | 4.76 | 21.31 | 13.957 | 2.933 | 0.210 | | | |
| As (mg kg ⁻¹) | 4.72 | 12.96 | 9.053 | 1.627 | 0.180 | | | |
| Cr (mg kg ⁻¹) | 5.96 | 24.10 | 14.298 | 4.060 | 0.284 | | | |

1.2 数据预处理

采集土壤光谱过程中往往会受到外界环境因素 与水分影响,导致所采集的光谱信息会产生较多的 噪声进而影响数据准确率,对其中受影响较大的光 谱数据进行剔除,再对剔除后的光谱数据进行预处 理。首先,通过对原始光谱反射率(R)进行 Savitzky-Golay(SG)平滑处理,提高光谱的平滑性,并降低 噪音的干扰,突出土壤光谱的特征^[15]。由于光谱数据 维数较高,需对其进行重采样来降低数据冗余,本 文选取 10 nm 重采样。在重采样的基础上,对原始 光谱数据进行几种常见光谱变换,如一阶微分 (FD)、二阶微分(SD)、倒数对数(LR)和连续 统去除(CR)。通过导数、倒数和微分处理可以进 一步减少噪声对光谱数据的影响,提升光谱数据与 土壤重金属的敏感性^[16],有利于光谱数据与土壤重金 属的相关性分析^[17]。

1.3 模型方法与理论

1.3.1 支持向量机 支持向量机是 Vapnik 等人基于 统计学习理论提出的一种机器学习方法。其基本思 想是通过支持向量机建立一个超平面作为决策曲面, 对不同类别的样本进行分割,分割原则为间隔最大 化。设一个样本集为 (x_i, y_i) , i = 1, ..., m, $x_i \in R^n$, m 为样本个数, $y_i \in \{-1, 1\}$, x_i 为第 i 个的特征向量, y_i 为 x_i 对应的类标记,其分割超平面所表示的公 式为:

$$f(x) = \omega^T x + b \tag{1}$$

式中, ω 是权重向量, b 为截距向量, f(x) 为超平 面, 其目标函数为:

$$\min_{\substack{\omega,b\\\omega,b}} \frac{1}{2} ||\omega||^2$$

$$s.t.y_i(\omega^T x + b) \ge 1, i = 1, 2, \cdots, m$$
(2)

对于线性不可分问题需要将样本映射到高维空

间,能够在高维空间的将其转化为线性可分问题, 令 $\phi(x)$ 表示映射后的特征向量,引人这种映射后对其 所求的对偶问题求解,则不需求解出真正的映射函 数,而要知道其核函数。

$$k(x_i, x_j) = \phi(x_i)^T \phi(x_j) \tag{3}$$

引入核函数之后,其函数表示为:

$$f(x) = \sum_{i=1}^{m} \alpha_i y_i k(x, x_i) + b \tag{4}$$

其中, α为拉格朗日因子, 根据土壤重金属反演的相 关特征,本文选取高斯径向基核函数对实验进行 研究。

1.3.2 遗传算法优化支持向量机 使用支持向量机 做回归分析时,其性能的效果主要受参数的影响。本文将选取遗传算法对参数进行优化,通过对初始 种群筛选进化,最终获取一个最适应环境的种群,从而获取最优解。遗传算法的参数设置,设置惩罚 因子 *C* 和核函数 *g* 的取值范围在 0~100 之间,初始 参数的进化代数为 1,种群数量为 20,迭代数为 200。适应度函数值使用均方根误差(*MSE*)进行计算:

$$fit_{MSE} = \sqrt{\frac{\sum_{i=1}^{n} (y_i - f_i)^2}{n}}$$
(5)

其中, fit为适应度函数, y_i 为样本真值, f_i 为预测值, n为样本数量。

2 结果与分析

2.1 相关性分析

本次试验主要针对 Fe、As 与 Cr 三种土壤重金 属进行相关研究,使用 SPSS 软件分别对 R、FD、 SD、LR 和 CR 等光谱数据与三种土壤重金属含量的 样本实测值进行相关性分析。从图 1 可以看出,土 壤重金属与 LR 和 CR 的相关性不明显,在某些波段 甚至低于 *R* 与土壤重金属的相关性。经过微分变换 后的光谱数据与土壤重金属的相关性得到了明显的 提升,该结果与徐明星等研究一致^[18],其中 SD 光谱 变换与土壤重金属 Fe、As 与 Cr 的相关性最佳,最 大相关系数分别为 0.508、0.518 与 0.545。因此,本 文选取二阶微分光谱变换数据对后续实验进行 研究。

2.2 特征提取

由于光谱数据维数大且存在数据冗余等因素,

为避免数据冗余并提高建模精度,本文将对光谱数 据进行皮尔森相关系数特征波段提取与主成分分析 特征提取。

2.2.1 皮尔森相关系数 (Pearson correlation coefficient, PCC) 可以更好的发现土壤重金属与光谱反射率的 敏感度,通过相关性分析提取特征光谱波段用于后 续反演土壤重金属含量研究。选取 SD 变换光谱与重 金属含量进行皮尔森相关性分析,当显著性水平 P=0.05 和 P=0.01 时,其相关系数分别呈显著相关和 极显著相关。本文通过选取显著性水平 P<0.01 的光 谱数据作为特征波段。Fe、As 和 Cr 分别提取了 17、8 和 9 个特征波段 (表 2),其中 490 nm、500 nm、510 nm 和 530 nm 为三种重金属的共有特征波段。Cr 的特征波段在可见光波段,As 的特征波段主要在 可见光波段,Fe 的特征波段主要在可见光和近红外 波段。

2.2.2 主成分分析 (Principal component analysis, PCA) 是一种常用的数据降维算法,其主要思想是将原有的高维变量重新组合成一组线性无关的综合变量,从而获取更加少量的新变量替代尽可能多的原变量,即为所提取的主成分。主成分个数可根据实际情况进行选取,主要通过特征值大于1且累计贡献率达85%确定主成分个数,以保证尽可能多的包含原有变量的信息。通过对10 nm 重采样后的SD 光谱变换数据进行主成分分析选取特征值大于1 的主成分作为本研究的特征提取,其中提取的12 个主成分累计贡献率为94.221%。表3 描述了主成分分析下的特征值、贡献率与累计贡献率。

2.3 模型精度比较

本次试验选取主成分分析与相关性分析法提取 的特征变量作为光谱参量,分别建立 SVM 与 GA-SVM 土壤重金属含量反演模型。采用模型决定系数 (*R*²)和模型均方根误差(*RMSE*)对预测模型进行 精度评估。其建模结果如表 4 和表 5 所示,由于篇 幅有限仅展示重金属 As 在主成分分析和皮尔森相关 系数特征提取下的 GA-SVM 模型验证集的实测值与 预测值散点图(图 2,图 3)。

2.3.1 SVM 回归模型估测 本次实验共采集 58 个 样本点,土壤重金属 Fe、As 和 Cr 分别使用 58、58 和 54 个样本进行模型分析,其中 Fe 和 As 选取 40 个样本建模,18 个样本验证,Cr 选取 38 个样本建 模,16 个样本验证。采用两种特征提取下的光谱数



注:图中表示原始光谱数据及四种光谱变换数据分别与土壤重金属 Fe、Cr 和 As 元素间的相关系数折线图

图 1 不同光谱数据与土壤重金属相关性

Fig.1 Correlation between different spectral data and soil heavy metals

| 表 2 | 皮尔森相关系数特征波段提取 |
|-----|---------------|
| 表 2 | 皮尔森相天杀敛符征波段提取 |

| | | Table 2 Pearson correlation coefficient feature band extraction | | | | | |
|--------------------|---------------------------------|---|--|--|--|--|--|
| 重金属 Heavy metal | 特征波段数 Feature band number | 特征波段(nm) Characteristic band | | | | | |
| Fe | 17 | 460、490、500、510、530、570、580、590、740、750、770、790、890、1700、2200、2220、2270 | | | | | |
| As | 8 | 440、490、500、510、530、540、1010、2120 | | | | | |
| Cr | 9 | 460、490、500、510、530、540、570、580、590 | | | | | |

表 3 SD 变换光谱主成分特征值、贡献率与累计贡献率

Table 3 Principal component characteristic value, contribution rate and cumulative contribution rate

| PC_1 | PC ₂ | PC ₃ | PC_4 | PC_5 | PC_6 | PC_7 | PC_8 | PC ₉ | PC ₁₀ | PC ₁₁ | PC ₁₂ |
|--------|---|---|---|--|---|--|---|--|---|---|---|
| 44.871 | 16.754 | 11.059 | 8.885 | 5.257 | 3.225 | 2.842 | 2.551 | 2.187 | 1.755 | 1.289 | 1.086 |
| 41.547 | 15.513 | 10.240 | 8.227 | 4.867 | 2.986 | 2.632 | 2.362 | 2.025 | 1.625 | 1.193 | 1.005 |
| 41.547 | 57.06 | 67.299 | 75.526 | 80.393 | 83.379 | 86.011 | 88.373 | 90.398 | 92.023 | 93.216 | 94.221 |
| | PC ₁ 44.871 41.547 41.547 | PC1 PC2 44.871 16.754 41.547 15.513 41.547 57.06 41.547 57.06 | PC1 PC2 PC3 44.871 16.754 11.059 41.547 15.513 10.240 41.547 57.06 67.299 | PC1 PC2 PC3 PC4 44.871 16.754 11.059 8.885 41.547 15.513 10.240 8.227 41.547 57.06 67.299 75.526 | $\begin{array}{c ccccccccccccccccccccccccccccccccccc$ | PC_1 PC_2 PC_3 PC_4 PC_5 PC_6 44.871 16.754 11.059 8.885 5.257 3.225 41.547 15.513 10.240 8.227 4.867 2.986 41.547 57.06 67.299 75.526 80.393 83.379 | $\begin{array}{c ccccccccccccccccccccccccccccccccccc$ | PC_1 PC_2 PC_3 PC_4 PC_5 PC_6 PC_7 PC_8 44.871 16.754 11.059 8.885 5.257 3.225 2.842 2.551 41.547 15.513 10.240 8.227 4.867 2.986 2.632 2.362 41.547 57.06 67.299 75.526 80.393 83.379 86.011 88.373 | $\begin{array}{c ccccccccccccccccccccccccccccccccccc$ | $\begin{array}{c ccccccccccccccccccccccccccccccccccc$ | $\begin{array}{c ccccccccccccccccccccccccccccccccccc$ |

注: SD为二阶微分光谱变换数据; PC_i(i=1, 2, 3…12)表示主成分数。

| 表 4 SVM 回归模型佰测 Table 4 SVM regression model estimation | | | | | | | | | |
|---|--|-------|--|-------|--|-------|--|-------|--|
| 香今届 | | PCA- | SVM | | | PCC- | SVM | | |
| 重亚声 Heavy metal | 建模集 <i>R</i> ² Calibration set <i>R</i> ² | RMSE | 验证集R ² Validation set R ² | RMSE | 建模集 <i>R</i> ² Calibration set <i>R</i> ² | RMSE | 验证集R ² Validation set R ² | RMSE | |
| Fe | 0.766 | 0.090 | 0.737 | 0.096 | 0.864 | 0.070 | 0.831 | 0.079 | |
| As | 0.716 | 0.117 | 0.447 | 0.128 | 0.668 | 0.127 | 0.469 | 0.128 | |
| Cr | 0.764 | 0.123 | 0.713 | 0.113 | 0.691 | 0.133 | 0.533 | 0.135 | |

| | 表 5 | GA-SVM 回归模型估测 |
|---------|-----|----------------------------------|
| Fable 5 | GA | -SVM regression model estimation |

| 重金属 – Heavy metal | | PCA-G | A-SVM | | PCC-GA-SVM | | | |
|----------------------|--|-------|---|-------|---|-------|--|-------|
| | 建模集 <i>R</i> ² Calibration set <i>R</i> ² | RMSE | 验证集 <i>R</i> ² Validation set <i>R</i> ² | RMSE | 建模集R ² Calibration set R ² | RMSE | 验证集R ² Validation set R ² | RMSE |
| Fe | 0.951 | 0.057 | 0.963 | 0.049 | 0.964 | 0.045 | 0.968 | 0.046 |
| As | 0.884 | 0.092 | 0.821 | 0.084 | 0.772 | 0.102 | 0.638 | 0.105 |
| Cr | 0.982 | 0.047 | 0.976 | 0.049 | 0.824 | 0.102 | 0.6 | 0.122 |





图 2 PCA-GA-SVM 土壤 As 含量散点图



实测值 (g kg⁻¹)

Measured value

据作为自变量建立反演重金属含量的 SVM 回归模型, 根据决定系数 R² 最大和均方根误差 RMSE 最小的原 则选择最佳模型,从表4中分析可知,重金属Fe、 As 和 Cr 的建模集 R² 分别为 0.864、0.716 和 0.764, RMSE 分别为 0.07、0.117 和 0.123, 模型估算预测效 果最好; 其模型验证 R²分别为 0.831、0.469 和 0.713, RMSE 分别为 0.079、0.128 和 0.113。通过分 析可知, 重金属 Fe 和 As 在皮尔森特征提取下的预

测精度较好, 而 Cr 在主成分特征提取下的预测精度 更高。

2.3.2 GA-SVM 回归模型估测 通过对二阶微分变 换光谱进行主成分特征提取和皮尔森特征波段提取, 使用两种特征波段的光谱数据分别作为自变量,重 金属含量作为因变量,用与 SVM 回归模型相同的建 模样本构建 GA-SVM 回归模型(表 5)。重金属 As 和 Cr 在主成分特征提取下建模集最优,其 R^2 分别 为 0.884 和 0.982, RMSE 分别为 0.092 和 0.047。而 Fe 在皮尔森特征提取下建模集最优,其 R² 为 0.964, RMSE为 0.045。相较于 SVM 回归模型 As 和 Cr 含 量预测 R² 分别提高了 0.168 和 0.218, RMSE 分别降 低了 0.025 和 0.076, 重金属 Fe 其 R² 提高了 0.1, RMSE 降低了 0.025。此外, 重金属 Fe、As 和 Cr 的 最优模型验证 R² 分别为 0.968、0.821 和 0.976,相 较于 SVM 回归模型其预测精度与模型稳定性都具有 明显提升。

讨论 3

本次试验结果表明,研究区土壤重金属含量与 原始光谱反射率相关性较低。由于原始土壤光谱数 据受环境等因素影响,存在一些噪声信息,且拥有 大量的波段信息, 使得土壤重金属光谱特征微弱, 敏感波段不易确定,进而导致土壤重金属与原始光 谱反射率相关性不明显。针对土壤重金属光谱特征 弱且敏感波段难以通过经验直接选取的问题,已有 研究表明利用光谱预处理结合特征提取能有效获取 土壤重金属敏感特征[19-20],本研究通过对原始光谱数 据进行平滑、重采样、光谱变换和特征提取等步骤,

增强土壤重金属与光谱数据敏感性,减少光谱数据 冗余,降低模型复杂度。此外,土壤重金属含量与 光谱数据间关系相对复杂,运用 SVM 可较好地解决 变量和因变量之间复杂的非线性关系,并利用 GA 算 法对 SVM 的参数进行优化,进一步提高模型反演精 度。与传统土壤重金属含量监测化学分析法对比, 利用高光谱遥感技术反演土壤重金属含量具有高效、 便捷、快速等优点,对土壤环境的动态监测和生态 修复提供数据支持具有重要意义。

研究可知,通过对光谱数据进行预处理,得到 二阶微分变换光谱与土壤重金属相关性整体最优, 和章琼等的研究基本一致^[21]。对重金属 Fe、As 和 Cr 筛选的敏感波段分析,在可见光波段 490 nm、 500 nm、510 nm 和 530 nm 处具有共同敏感特征,说 明各重金属间可能存在明显的相关关系,可在后续 研究进一步分析。对比两种反演模型,GA-SVM 模 型明显优于 SVM 模型,其决定系数 *R*² 普遍提高 0.1 以上,表明优化后的参数能够稳定提高模型反演精 度,可在一定精度条件下实现土壤重金属含量的快 速监测。本次研究主要针对耕地土壤重金属提出了 一种提高模型反演精度的方法,其是否适用于其他 土壤类型的研究,可在后续研究中进一步探索。

4 结论

(1)对原始光谱数据采用多种光谱变换,并与 土壤重金属进行相关性分析,得出二阶微分变换光 谱与土壤重金属 Fe、As 和 Cr 含量相关性最佳,其 最大相关系数分别为 0.508、0.518 与 0.545。

(2)通过 PCC 特征波段提取可得,重金属 Fe、 As 和 Cr 在 490 nm、500 nm、510 nm 和 530 nm 具 有共同特征波段。此外,对比两种不同特征波段提 取下的建模结果发现,重金属 As 和 Cr 在主成分分 析下的特征提取建模效果总体较好,其中 Fe 在皮尔 森相关性下的特征提取建模精度略优。

(3) 通过 SVM 与 GA-SVM 两种模型建模对比 分析可得, GA-SVM 回归模型 R²比 SVM 模型普遍 提高 0.1 以上,其 RMSE 也相应降低。表明通过遗传 算法优化后的支持向量机模型精度得到明显提高, 研究可为耕地土壤重金属含量高光谱反演提供参考。

参考文献:

[1] 郭云开,周烽松,丁美青,等.水稻冠层与土壤高光谱反演土壤
 重金属对比研究[J]. 遥感信息, 2017, 32(2): 173 - 179.

- [2] 张东辉, 赵英俊, 陆冬华, 等. 高光谱在土壤重金属信息提取中的应用与实现[J]. 土壤通报, 2018, 49(1): 31 37.
- [3] 李媛媛,李 微,刘 远,等.基于高光谱遥感土壤有机质含量
 预测研究[J].土壤通报,2014,45(6):1313-1318.
- [4] 黄长平,刘 波,张 霞,等. 土壤重金属Cu含量遥感反演的波 段选择与最佳光谱分辨率研究[J]. 遥感技术与应用, 2010, 25(3): 353-357+341.
- [5] Choe E, Meerf V D, Ruiteenbeek F V, et al. Mapping of heavy metal pollution in stream sediments using combined geochemistry, field spectroscopy, and hyperspectral remote sensing: a case study of the Rodalquilar mining area[J]. Remote Sensing of Environment, 2008, 112(7): 3222 – 3233.
- [6] 郭云开, 钱 佳, 蒋 明, 等. 基于核偏最小二乘的矿区土壤Cu 含量高光谱反演[J]. 土壤通报, 2019, 50(1): 52 56.
- [7] 龚绍琦,王 鑫,沈润平,等. 滨海盐土重金属含量高光谱遥感 研究[J]. 遥感技术与应用, 2010, 25(2): 169 - 177.
- [8] 涂宇龙, 邹 滨, 姜晓璐, 等. 矿区土壤Cu含量高光谱反演建 模[J]. 光谱学与光谱分析, 2018, 38(2): 575 - 581.
- [9] Vapnik V. The nature of statistical learning theory[M]. New York: Springer-Verlag, 1995: 138 - 145.
- [10] 王晓玲, 杜培军, 谭 琨. 高光谱遥感影像SVM分类中训练样本选择的研究[J]. 测绘科学, 2011, 36(3): 127 129.
- [11] 鄢文苗,任 东,黄应平,等.基于SVM土壤重金属污染评价的 训练数据集构建[J].武汉大学学报(理学版), 2019, 65(3): 316-322.
- [12] 沈 强,张世文,夏沙沙,等.基于支持向量机的土壤有机质高 光谱反演[J].安徽理工大学学报(自然科学版),2019,39(4):39-45.
- [13] 吕杰,刘湘南.利用支持向量机构建水稻镉含量高光谱预测 模型[J].应用科学学报,2012,30(1):105-110.
- [14] Wang X J, Zhang Z P. A comparison of conditional simulation, Kriging and trend surface analysis for soil heavy metal pollution pattern analysis[J]. Journal of Environmental Science & Health Part A, 1999, 34(1): 73 – 89.
- [15] Xu M X, Wu S H, Zhou S L, et al. Hyperspectral reflectance models for retrieving heavy metal content: application in the archaeological soil[J]. Journal of Infrared and Millimeter Waves, 2011, 30(2): 109 – 114.
- [16] 王 璐, 蔺启忠, 贾 东, 等. 基于反射光谱预测土壤重金属元 素含量的研究[J]. 遥感学报, 2007, 25(6): 906-913.
- [17] 陈红艳, 赵庚星, 张晓辉, 等. 去除水分影响提高土壤有机质含 量高光谱估测精度[J]. 农业工程学报, 2014, 30(8): 91-100.
- [18] 徐明星,吴绍华,周生路,等.重金属含量的高光谱建模反演:考 古土壤中的应用[J].红外与毫米波学报,2011,30(2):109-114.
- [19] 滕 靖,何政伟,倪忠云,等.西范坪矿区土壤铜元素的高光谱 响应与反演模型研究[J].光谱学与光谱分析,2016,36(11): 3637-3642.
- [20] 古力扎提·艾买提,阿不都拉·阿不力孜,茹克亚·沙吾提,等.准 东煤田土壤铅含量高光谱估算[J]. 土壤通报, 2018, 49(5): 1233-1239.
- [21] 章 琼, 郭云开, 钱 佳, 等. 像元二分模型的土壤重金属Cr的 遥感反演[J]. 测绘科学, 2020, 45(4): 58-63.

The Hyperspectral Inversion Method of Heavy Metal Contents in Cultivated Soils Based on GA-SVM

GUO Yun-kai^{1,2}, ZHANG Si-ai^{1,2*}, XIE Xiao-feng^{1,2}, XIE Qiong^{3,4}

 (1. School of Traffic and Transportation Engineering, Changsha University of Science and Technology, Changsha 410076, China;
 2. Institute of Surveying and Mapping Remote Sensing Applied Technology, Changsha University of Science and Technology, Changsha 410076, China;
 3. Department of Surveying and Mapping Geography, Hunan vocational college engineering, Changsha 410151, China;
 4. Engineering Laboratory of Spatial Information Technology of Highway Geological Disaster Early Warning, Changsha University of Science and Technology, Changsha 410114, China)

Abstract: In order to improve the accuracy of hyperspectral inversion model for heavy metal content in cultivated soil, a genetic algorithm (GA) optimized support vector machine (SVM) was proposed to retrieve the heavy metal content of cultivated soil in a certain area of Yueyang County. After SG smoothing and 10 nm resampling, the first-order / second-order differential, reciprocal logarithm and continuum removal spectral transformation methods were used to enhance the spectral characteristics. The optimal transform spectra were selected by correlation analysis. Pearson correlation coefficient and principal component analysis were used to extract the spectral characteristic variables of heavy metals. SVM and GA-SVM were used to establish soil heavy metal hyperspectral inversion models and their accuracies were verified. The results showed that the correlation between the second-order differential transform spectra and the contents of heavy metals was the most prominent. The visible light bands of 490 nm, 500 nm, 510 nm and 530 nm were the most prominent compared with the SVM regression model. The prediction accuracy was significantly improved, and the verification set R² values of Fe, As and Cr were 0.968, 0.821 and 0.976, respectively. The research results could provide a new reference for the application of remote sensing technology to retrieve the content of heavy metals in cultivated soil.

Key words: Soil heavy metal; Hyperspectral remote sensing; Spectral transformation; Genetic algorithm; Support vector machine

[责任编辑:韩春兰]