

可迁移的土壤重金属污染高光谱定性分类方法研究

陶超¹, 崔文博¹, 王亚晋¹, 邹滨^{1,2*}, 邹嵘嵘¹

1. 中南大学有色金属成矿预测与地质环境监测教育部重点实验室, 地球科学与信息物理学院, 湖南长沙 410083
2. 国家重金属污染防治工程技术研究中心, 湖南长沙 410083

摘要 现有基于高光谱遥感的土壤重金属污染定性分类模型, 大多采用同一地区室内光谱测定训练样本数据进行模型构建与测试。但室内光谱测定需要复杂的处理过程, 成本高, 效率低, 且无法快速获得目标区域空间上连续的光谱信息。考虑到实际应用需求, 模型在相同实验区和不同试验区野外光谱数据是否具有较好的迁移推广能力是目前迫切需要回答的问题。为回答这一问题, 选取湖南省郴州市和衡阳市两铅锌矿区作为实验研究区, 选用支持向量机(SVM)作为分类器, 将郴州实验区室内采样的83个样本数据和衡阳实验区室内采样的46个样本数据分别用于分类器训练, 将衡阳地区野外采样的46个样本数据用于分类测试。并首先通过基于联合分布适配(JDA)的迁移学习方法进行光谱变换以缩小两地室内外测定光谱分布差异, 然后进行不同区域室内外土壤重金属污染定性分类模型迁移。实验结果表明: (1)由于野外测得的光谱数据会受到太阳辐射、提取的土壤成分差异等因素的干扰导致室内外光谱数据存在显著的分布差异, 难以直接将基于室内采样数据训练得到的土壤重金属污染定性分类模型迁移到同一地区测定的野外高光谱数据上。但通过JDA变换缩小室内外分布差异后, 模型迁移能力得到显著提升, 砷(As)、铅(Pb)和锌(Zn)三种重金属含量是否超标的分类精度都达到了84%以上, Zn元素含量是否超标的分类精度甚至达到了89%以上。(2)由于季节性影响、地区成分的干扰和光谱噪声的增加, 不同地区光谱数据存在着更为显著的分布差异, 加大了不同地区土壤重金属污染监测的难度, 难以将基于室内采样光谱数据所建立的土壤重金属定性分类模型直接迁移到其他地区野外采样数据上(平均分类精度仅在50%左右)。经过JDA迁移学习方法进行室内外光谱变换处理后, 模型迁移能力得到保证, 因此, 室外光谱采样可直接用于研究不同试验区重金属(As, Pb和Zn)的污染情况。

关键词 高光谱遥感; 土壤重金属; 迁移学习; 室内外光谱采样

中图分类号: P237 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2019)08-2602-06

引言

随着我国现代工业化速度的加快和人类活动的影响, 土壤重金属富集愈发严重。砷(As)、铅(Pb)和锌(Zn)作为土壤三大主要污染元素, 具有污染时间长, 范围广的特点^[1], 且能够随食物链进入人体, 危害人类身体健康^[2], 快速探测土壤重金属含量并探究其污染分布成为当务之急^[3]。传统的化学方法成本高, 效率低, 难以对大面积区域土壤重金属含量进行实时监测。

高光谱遥感具有快速、低成本、无污染的特点^[4], 加之分辨率高, 波段连续性强的优势, 已被广泛用于土壤物质含

量的估算。近年来, 国内外研究人员已相继利用高光谱数据对土壤重金属含量进行了反演研究, 如: Sun^[5]等利用光谱聚类, 将遗传算法和偏最小二乘回归法相结合, 建立土壤重金属含量的预测模型; Dong^[6]等采用多变量线性回归等方法, 对土壤重金属含量和光谱数据进行了定量建模; Shi^[7]等采用人工神经网络等方法建立模型。但是, 目前利用高光谱数据进行土壤重金属反演仍存在以下两个主要问题:

(1)太阳辐射等环境因素干扰研究所用光谱数据大多依赖于室内采集, 而室内光谱测定昂贵低效, 无法快速获得目标区域空间上连续的光谱信息^[3]。

(2)由于地区成分的干扰、光谱噪声的增加, 将建立的回归模型直接迁移到其他地区是有困难的^[2], 这进一步加大

收稿日期: 2018-06-08, 修订日期: 2018-10-16

基金项目: 国家自然科学基金项目(41771458), 湖南省自然科学基金项目(2017JJ3378), 国家重点研发项目(2018YFB0504500)资助

作者简介: 陶超, 1985年生, 中南大学地球科学与信息物理学院测绘与遥感科学系副教授 e-mail: kingtaochao@126.com

* 通讯联系人 e-mail: 210010@csu.edu.cn

了土壤重金属浓度监测的成本和代价。

为实现基于高光谱遥感低价高效的土壤重金属污染监测,陶超^[8]等比较了高光谱土壤重金属定量反演模型和定性分类模型在不同地域的可迁移能力,实验表明在快速检测土壤重金属污染的问题上,定性分类是一种更加切实可行的方式。但该工作仍使用室内测定的光谱数据。已训练好的定性分类模型是否可以迁移到野外测定的光谱数据?是否可以联合室内外高光谱提高模型分类精度?为回答这些问题,选取湖南省郴州市和衡阳市两铅锌矿区作为研究区,首先通过联合分布适配(joint distribution adaptation, JDA)的迁移学习方法缩小两地室内外光谱分布差异,并进行不同域室内外土壤重金属污染定性分类模型迁移。实验结果表明:由于相同和不同地区室内外采样的光谱数据分布差异较大,难以将基于室内采样的光谱数据建立的重金属污染定性分类模型应用于同一或其他地区野外光谱数据。但通过 JDA 方法缩小两地室内外光谱分布差异后,模型的迁移能力得到有效提高。

1 实验部分

1.1 研究区概况

选取湖南省郴州市(113.02E, 25.46N)和衡阳市(112.39E, 26.98N)两铅锌矿区作为研究区,以郴州某矿区采集的 83 个样本数据和衡阳某矿区采集的 46 个样本数据作为实验数据。研究区属于长江中游地区,气候温暖,四季分明,土壤以砖红壤为主。研究区地理位置基本情况及采样点分布如图 1 所示。

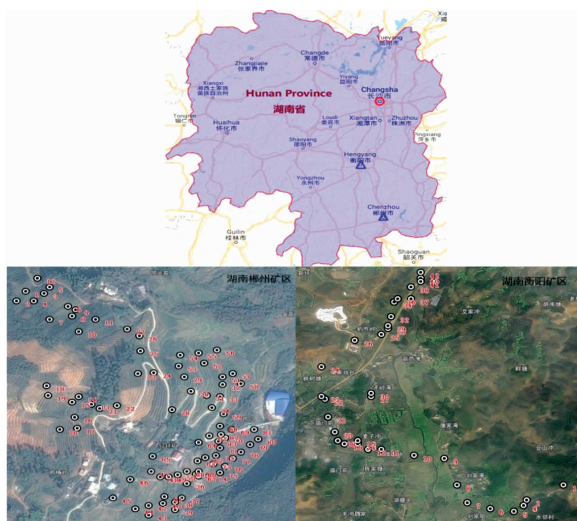


图 1 研究区地理位置和采样点分布

Fig. 1 Geographical location of the study area and distribution of sampling points

1.2 研究方法

1.2.1 土壤重金属浓度测定和高光谱数据采集

采样点均选取 10 m² 范围五点采样法采集表层 30 cm 内的土壤。实验室内数据采集后需将土壤放置在室内风干处理,并对土壤进行研磨,过 100 目土壤筛,测定郴州和衡阳

土壤中 As, Pb 和 Zn 三种重金属的浓度。

野外光谱数据采集用到的仪器为 PSR-3500 野外便携式地物光谱仪器,测量时,探头方向与太阳入射角方向相对。室内光谱采集需利用同一台 PSR-3500 野外便携式地物光谱仪器进行光谱测定,采集过程中需要使用专用的灯光作为光源进行量测,并且衡阳地区的室内光谱采样与室外光谱采样选用同一批土样,在测量一定数量样本后,需利用参考白板进行辐射校正,并将数据进行存储。

1.2.2 光谱数据预处理

实验共采集 1 024 个波段的光谱数据,利用重采样的预处理方法,选定 10 nm 间隔求均值的方法,处理完成后得到 208 个波段的光谱数据。

郴州和衡阳两个地区室内光谱特性大体相似^[9],略有不同,而衡阳野外测量的光谱数据由于缺少室内处理过程,与室内采样光谱数据相比具有显著分布差异^[10],所以难以直接将室内采样光谱数据建立的重金属污染定性分类模型应用于同一或不同区野外光谱数据上(见表 3 和表 4)。为解决上述问题,选用 JDA 的迁移学习方法减小室内外光谱数据的分布差异以提高模型迁移分类能力。

1.2.3 联合分布适配方法

(1)联合分布适配方法原理

选用 JDA 的迁移学习方法,力图寻找变换矩阵 E ,使变换后室内外光谱的边缘分布和条件分布距离足够小^[11]。该方法所用符号和具体过程如表 1 所示。

表 1 本方法使用的符号和描述

Table 1 The symbols and descriptions used in this method

符号	描述	符号	描述
K_s, K_t	室内/室外光谱数据集	M_c	MMD 矩阵, $c=0, \dots, C$
p, q	室内/室外样本个数	E	变换矩阵
D	MMD 距离	c	样本标签
X	输入的数据矩阵	H	中心矩阵
λ	正则参数	I	单位矩阵

1)适配边缘分布(使边缘概率 $P(E^T x_s)$ 和 $P(E^T x_t)$ 的距离最小): 设 x_s 和 x_t 分别为室内光谱和室外光谱的数据矩阵,室内光谱和室外光谱的最大均值差异可用式(1)的最大均值差异(MMD)距离来表示

$$D(K_s, K_t) = \left\| \frac{1}{p} \sum_{s=1}^p E^T X_s - \frac{1}{q} \sum_{t=1}^q E^T X_t \right\|^2 \quad (1)$$

2)适配条件分布(使条件概率 $P(y_s | E^T x_s)$ 和 $P(y_t | E^T x_t)$ 距离最小): 光谱数据中类与类之间的 MMD 距离可以根据式(2)

$$D(K_s, K_t) = \sum_c \left\| \frac{1}{p_c} \sum_{x_s \in K_s^{(c)}} E^T x_s - \frac{1}{q_c} \sum_{x_t \in K_t^{(c)}} E^T x_t \right\|^2 \quad (2)$$

3)最后将边缘分布和条件分布两种适配方式统一起来, $\lambda \|E\|^2$ 作为正则项,并避免出现 $E=0$ 的情况,得到总的优化目标为

$$\min \left\| \frac{1}{p} \sum_{s=1}^p \mathbf{E}^T \mathbf{X}_s - \frac{1}{q} \sum_{t=1}^q \mathbf{E}^T \mathbf{X}_t \right\|^2 + \sum_c \left\| \frac{1}{p_c} \sum_{X_s \in K_s^{(c)}} \mathbf{E}^T \mathbf{X}_s - \frac{1}{q_c} \sum_{X_t \in K_t^{(c)}} \mathbf{E}^T \mathbf{X}_t \right\|^2 + \lambda \|\mathbf{E}\|^2 \quad (3)$$

$$\mathbf{E}^T \mathbf{X} \mathbf{H} \mathbf{X}^T \mathbf{E} = \mathbf{I}$$

(2) JDA 变换前后的光谱数据差异可视化对比分析
 为验证 JDA 变换能否有效缩小室内外光谱数据分布差

异, 在衡阳地区受污染和不受污染样本中各选一个样本数据, 同时在郴州地区选出重金属含量最为接近的对应样本, 可视化它们在 JDA 变换前后的光谱曲线, 由于 400 nm 以下和 2 400 nm 以上的光谱的信噪比过低, 且 1 360~1 490 和 1 810~1 960 nm 的波段存在水汽吸收的干扰, 所以这些波段需要被剔除, 结果如图 2 和图 3 所示。

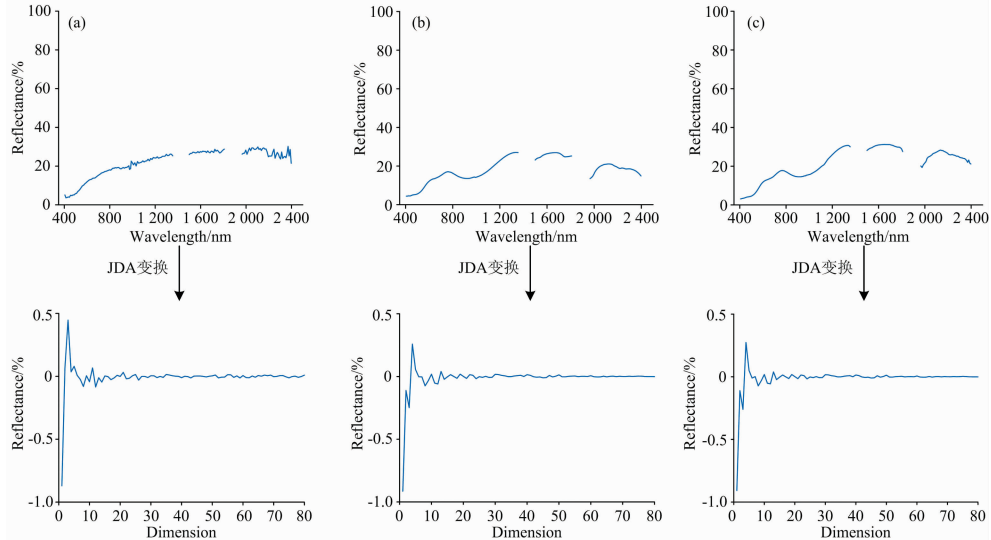


图 2 郴州室内和衡阳室内外受污染样本光谱图像

(a): JDA 前后郴州室内样本光谱曲线; (b): JDA 前后衡阳室内样本光谱曲线; (c): JDA 前后衡阳室外样本光谱曲线

Fig. 2 Spectral reflectance of contaminated indoor and outdoor samples from Chenzhou and Hengyang

(a): Spectral curves of Chenzhou indoor samples before and after JDA; (b): Spectral curves of Hengyang indoor samples before and after JDA; (c): Spectral curves of Hengyang outdoor samples before and after JDA

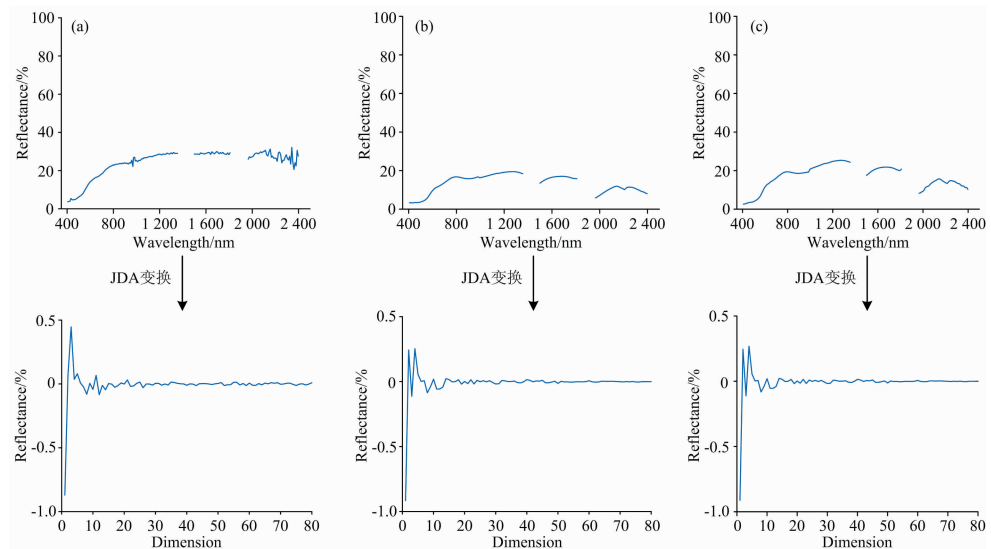


图 3 郴州室内和衡阳室内外未受污染样本光谱图像

(a): JDA 前后郴州室内样本光谱曲线; (b): JDA 前后衡阳室内样本光谱曲线; (c): JDA 前后衡阳室外样本光谱曲线

Fig. 3 Spectral reflectance of uncontaminated indoor and outdoor samples from Chenzhou and Hengyang

(a): Spectral curves of Chenzhou indoor samples before and after JDA; (b): Spectral curves of Hengyang indoor samples before and after JDA; (c): Spectral curves of Hengyang outdoor samples before and after JDA

从图中可以看出,受污染样本与未受污染样本 JDA 前后的光谱曲线大体相似,细微之处略有不同。且对于这两类样本, JDA 前不同地区采样数据(特别是室内外采样数据)的光谱差异非常明显。但经过 JDA 变换后,两者差异均明显减小,这为后续的模型能够有效迁移提供了保证。

2 结果与讨论

2.1 研究区土壤污染分类

根据土壤环境质量标准(GB15618—1995),对郴州和衡阳研究区的重金属含量数据进行分析,将样本分为受污染和不受污染两个类别,其中含量是否超过国家二级标准含量作为是否受污染的判断依据,两个样本区污染情况如表 2 所示。

2.2 实验与结果

实验选用支持向量机(support vector machine, SVM)作为分类器,将郴州实验区室内采样的 83 个样本数据和衡阳实验区室内采样的 46 个样本数据分别用于分类器训练,将

衡阳地区野外采样的 46 个样本数据用于分类测试,进行如下两个实验:

实验一:首先利用衡阳某矿区 JDA 变换前后室内采样土壤高光谱数据作为训练样本数据训练土壤重金属污染定性分类模型,然后分别输入同地区 JDA 变换前后野外采样高光谱数据作为测试样本数据进行测试,实验结果如表 3 所示。

表 2 研究区土壤重金属的污染情况分布

Table 2 The distribution of soil heavy metal pollution in research areas

	受污染样本数	未受污染样本数
郴州(As)	54	29
郴州(Pb)	43	40
郴州(Zn)	50	32
衡阳(As)	36	10
衡阳(Pb)	36	10
衡阳(Zn)	38	8

表 3 衡阳室内数据训练的分类模型用于室外采样数据的分类结果

Table 3 Classification results of outdoor sampling data using indoor data trained model in Hengyang area

重金属类别	训练样本 (衡阳试验区室内采样光谱数据)	测试样本 (衡阳试验区野外采样光谱数据)	SVM 直接 分类精度/%	JDA 光谱变换后的 SVM 分类精度/%
As	46	46	78.26	84.78
Pb	46	46	78.26	86.96
Zn	46	46	86.96	89.14

从表 3 可以看出:同一地区基于室内采样光谱数据的分类模型对基于野外采样的光谱数据具有一定的可迁移能力,直接分类后三种元素含量是否超标的平均分类精度在 80% 左右。在利用 JDA 的方法对基于室内外采样的光谱数据缩小分布差异后,模型的迁移能力进一步得到提高,平均分类

精度达到了 86% 以上。

实验二:利用郴州某矿区 JDA 变换前后室内采样光谱数据作为训练数据训练土壤重金属污染定性分类模型,然后分别输入衡阳地区 JDA 变换前后野外采样光谱数据作为测试数据进行模型测试。实验结果如表 4 所示。

表 4 郴州室内数据训练的分类模型用于衡阳地区室外采样数据的分类结果

Table 4 Classification resultsof outdoor sampling data in Hengyang area using indoor data trained model in Chenzhou area

重金属类别	训练样本 (郴州试验区室内采样光谱数据)	测试样本 (衡阳试验区野外采样光谱数据)	SVM 直接 分类精度/%	JDA 光谱变换后的 SVM 分类精度/%
As	83	46	76.09	84.78
Pb	83	46	34.78	80.43
Zn	82	46	52.17	86.96

从表 4 可以看出:基于室内光谱数据训练的分类模型难以直接迁移到不同地区室外样本光谱数据中。其中,土壤中 Pb 元素含量是否超标的分类精度仅为 34.78%,三种重金属的平均分类精度不到 55%,在利用 JDA 方法对不同地区光谱数据进行变换处理后,模型的迁移能力得到保证,平均分类精度达到 84% 以上。

3 结 论

以湖南省郴州市和衡阳市两大铅锌矿区作为研究区,以郴州市某矿区室内采样的高光谱数据和衡阳市某矿区室内外采样的高光谱数据作为实验数据,以 As, Pb 和 Zn 三种重金属元素含量是否超标作为分类对象,针对如何将基于室内采

样数据训练得到的土壤重金属污染定性分类模型迁移到同一或不同区域野外采样光谱数据的问题进行实验研究探索, 实验结论如下:

(1) 由于野外测得的光谱数据会受到太阳辐射、提取的土壤成分差异等因素的干扰导致室内外光谱数据存在显著分布差异, 加之季节性影响、地区成分的干扰造成的不同地区区间的光谱数据分布差异, 难以直接将室内训练得到的土壤

重金属高光谱污染定性分类模型迁移到同一地区和不同地区野外光谱数据上。

(2) 经过 JDA 迁移学习方法进行室内外光谱变换处理后, 模型迁移能力得到有效提高, 为实现基于高光谱遥感低成本、高效率的土壤重金属定性污染监测提供了一种切实可行的方法。

References

- [1] ZHANG Yin-ling, YAO Feng, NIU Ting, et al(张银玲, 姚峰, 牛婷, 等). Environmental Protection of Xinjiang(新疆环境保护), 2016, 38(3): 15.
- [2] Wang F, Gao J, Zha Y. ISPRS Journal of Photogrammetry & Remote Sensing, 2018, 136: 73.
- [3] Shi T, Chen Y, Liu Y, et al. Journal of Hazardous Materials, 2014, 265(2): 166.
- [4] XIAO Jie-ying, WANG Yan, ZHANG Qian, et al(肖捷颖, 王燕, 张倩, 等). Hubei Agricultural Sciences(湖北农业科学), 2013, 52(6): 1248.
- [5] Sun W, Zhang X, Zou B, et al. Remote Sensing, 2017, 9(6): 632.
- [6] Dong J, Dai W, Xu J, et al. International Journal of Environmental Research & Public Health, 2016, 13(7): 640.
- [7] Shi T, Liu H, Chen Y, et al. Sensors, 2017, 17(5): 1036.
- [8] TAO Chao, WANG Ya-jin, ZOU Bin, et al(陶超, 王亚晋, 邹滨, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2018, 38(6): 1850.
- [9] SHANG Xuan, LI Xi-can, XU You-you, et al(尚璇, 李西灿, 徐邮邮, 等). Scientia Agricultura Sinica(中国农业科学), 2017, 50(8): 1465.
- [10] Piekarczyk J, Kazmierowski C, Krolewicz S, et al. IEEE Journal of Selected Topics in Applied Earth Observations & Remote Sensing, 2016, 9(2): 827.
- [11] Long M, Wang J, Ding G, et al. Transfer Feature Learning with Joint Distribution Adaptation, 2013 IEEE International Conference on Computer Vision, 2013. 2200.

Soil Heavy Metal Qualitative Classification Model Based on Hyperspectral Measurements and Transfer Learning

TAO Chao¹, CUI Wen-bo¹, WANG Ya-jin¹, ZOU Bin^{1, 2*}, ZOU Zheng-rong¹

1. The Key Laboratory of Metallogenic Prediction of Nonferrous Metals and Geological Environment Monitoring (Center South University), Ministry of Education, School of Geoscience and Info-Physics, Changsha 410083, China

2. Chinese National Engineering Research Center for Control & Treatment of Heavy Metal Pollution, Changsha 410083, China

Abstract The current qualitative classification models of soil heavy metal content based hyperspectral remote sensing technology mostly use indoor measured spectral data from the same area for model training and testing. However, the indoor spectrum measurement requires a complicated processing process with high cost and low efficiency, and thus cannot obtain the spatially continuous spectral information in the target area quickly. Moreover, whether this kind of model can be transferred to the outdoor measured spectral data in different test areas is still unclear. In order to answer this question, two lead-zinc mining areas in Chenzhou City and Hengyang City of Hunan Province were selected as research areas. Support Vector Machine was used as classifier. Then 83 sample data from indoor sampling in Zhangzhou experimental area and 46 sample data from indoor sampling in Hengyang experimental area were used for classifier training, and 46 sample data from field sampling in Hengyang area were used for classification testing. The difference of spectral distribution between the indoor and outdoor measured spectral data was reduced by the transfer learning method based on joint distribution adaptation (JDA), and then the domain adaption model for two research areas was constructed. The experimental results show that: (1) The spectral data measured by outdoor samples may be affected by factors such as solar radiation and differences in extracted soil components, leading to the significantly spectral difference for indoor and outdoor samples. As a result, it is difficult to directly transfer the qualitative classification model of

soil heavy metal pollution trained by indoor samples to the outdoor samples from the same area. However, after the reduction of indoor and outdoor distribution differences by JDA transformation, the transfer ability of the model has been significantly improved, and the classification accuracy of three heavy metals As, Pb and Zn has reached over 84%. The accuracy of classification of Zn elements exceeding the standard even reached 89%. (2) Due to seasonal influences, regional component interference, and spectral noise, there are even more significant differences in the distribution of spectral data in different areas. This further increases the difficulty of soil heavy metal pollution monitoring in different areas, and it is difficult to directly transfer the qualitative classification model of soil heavy metals based on indoor sampling spectral data to field sampling data in other areas (with an average classification accuracy of about 50%). After the indoor and outdoor spectral transformation processing by JDA, the transfer ability of the model has been greatly improved. Therefore, the outdoor spectral sampled can be directly used to investigate the pollution situation of heavy metals (As, Pb and Zn) in different test areas.

Keywords Hyperspectral remote sensing; Heavy metal in soil; Transfer learning; Indoor and outdoor spectral sampling

(Received Jun. 8, 2018; accepted Oct. 16, 2018)

* Corresponding author