

高光谱的矿区植物异常信息提取

崔世超^{1, 2, 3}, 周可法^{1, 2*}, 丁汝福⁴

1. 中国科学院新疆生态与地理研究所新疆矿产资源研究中心, 新疆 乌鲁木齐 830011
2. 新疆矿产资源与数字地质重点实验室, 新疆 乌鲁木齐 830011
3. 中国科学院大学, 北京 100049
4. 有色金属矿产地质调查中心, 北京 100012

摘要 白茎绢蒿是一种广泛分布于新疆富蕴县各个矿区的一种植物。在矿区进行矿产勘查时, 由于植物等障碍信息存在, 传统的勘查方法已经难以发挥作用, 急需一些新方法、新思路。遥感植物地球化学方法可以巧妙地利用植物这一天然的信息源, 把植物从障碍信息转换为有用信息。帮助人们快速、经济地获取植物屏障下的矿产有用信息。由于其具有大面积、快速、无损性等优点, 受到了越来越多学者的关注, 成为当下的研究热点。近些年虽然有学者综合考虑“吸收系数”和“衬度系数”这两个指标, 证明了白茎绢蒿是对隐伏矿床的勘查具有较好指示性作用的植物, 生在在矿床上部的植物可以较好的吸收土壤中的成矿元素, 在其体内形成地球化学异常, 相比于其他植物异常信息更加清晰可见。但是目前没有人研究是否可以从光谱的角度来发现白茎绢蒿体内的地球化学异常, 进而为隐伏矿床的勘查提供参考。因此, 本研究首次尝试从白茎绢蒿的光谱信息中寻找出与地球化学异常密切相关的特征波段或者特征值, 然后构建基于植物光谱的隐伏矿床预测模型。采取的方法是首先利用 ASD FieldSpec3 型光谱仪分别对生长在矿床上部和背景区的植物进行光谱测定, 然后从原始光谱、一阶导数光谱、二阶导数光谱、一阶导数的分形维数、二阶导数的分形维数五个层面对生长在这两个区域的植物光谱进行对比分析, 最终优选出了 10 个差异显著的特征波段, 分别为: R'_{824} , R'_{834} , R'_{1533} , R'_{1573} , R'_{1633} , R'_{1643} , R'_{1284} , R'_{1703} , 一阶导数的分形维数以及二阶导数的分形维数。这些特征波段可以作为植物地区寻找隐伏矿床的植物地球化学标志。以优选出的 10 个特征波段作为输入参数, 分别用随机森林 (RF) 和偏最小二乘-支持向量机 (PLS-SVM) 构建了基于植物光谱数据的隐伏矿床预测模型。结果表明: (1) 两种模型均可以取得较好的效果, 但是相比于随机森林模型, 偏最小二乘-支持向量机模型具有更好的鲁棒性, 泛化能力也更强; (2) 利用植物的光谱异常寻找隐伏矿床具有较大的潜力, 因为相比于传统方法, 更加简单、快速。课题组已经利用动力三角翼和 HySpex 成像高光谱传感器构建了“超低空探测平台”, 可以实现对地“亚米级”的观测。但是如何有效的解决“空间尺度”和“光谱尺度”问题, 如何把地面试验场建立的模型更好的应用于超低空探测平台, 实现研究区大面积地、快速地植物异常信息提取将是我们的下一步的研究重点。

关键词 高光谱; 隐伏矿床; 偏最小二乘-支持向量机; 随机森林; 地球化学标志

中图分类号: P627 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2019)01-0241-09

引言

随着经济的快速发展, 社会对矿产资源的需求量越来越大。但是目前露头或者浅层的矿体日益减少, 全世界陆地的 2/3 被中等到茂密的植被覆盖, 是一个重要的找矿方向。如

果可以在此区域快速地进行矿产探查, 可以极大地缓解目前资源日益短缺的燃眉之急。已有大量的研究表明: 在深覆盖区域使用植物地球化学方法找矿是一种行之有效的方法, 其原理是生长在矿床上部的某些“指示性”植物可以较好地吸收成矿元素, 并在其体内形成地球化学异常, 可以根据这些异常判断隐伏矿床的大小和位置。相比于传统方法, 由于使用

收稿日期: 2017-12-11, 修订日期: 2018-05-05

基金项目: 国家自然科学基金(联合基金项目-重点支持项目)(U1503291), 新疆维吾尔自治区国际科技合作计划项目(20156017)资助

作者简介: 崔世超, 1991 年生, 中国科学院新疆生态与地理研究所博士研究生 e-mail: 1209048205@qq.com

* 通讯联系人 e-mail: zhoukf@ms.xjb.ac.cn

高光谱技术具有快速、省时、省力等优势,因此越来越多的学者尝试使用光谱技术来发现由于重金属胁迫造成的植物地球化学异常,并且提出了一系列参数用来表征植物受重金属的胁迫程度,这些参数主要包括以下三种:

(1) 植被指数:在研究早期,学者们主要是使用归一化植被指数(NDVI)、增强型植被指数(EVI)以及比值植被指数(RVI)等传统植被指数来表征植物受重金属的胁迫程度。如 Rathod 等^[1]研究表明,一些叶绿素指数和水分指数与植物中的砷含量具有较好的相关性,并且可以较好的识别出植物受金属的胁迫程度;Sridhar 等^[2]使用遥感光谱和 Landsat TM 影像分析几种重金属元素胁迫下大豆的生长状况,发现随着生物量的增加,大豆叶片和根系中的重金属含量也显著增多,并且归一化植被光谱参数(NDVI)可以用来监测大豆受重金属胁迫的程度,二者呈现显著的负相关关系;但是以上研究使用的植被指数都是已经存在的植被指数,这些植被指数最初设计的目的往往不是用来监测重金属胁迫的。设计一些专门表征重金属胁迫程度的植被指数是十分必要的。Hede 等^[3-6]相继提出一些新的可以定量地表征植物受重金属的胁迫程度的植被指数,并证明相比于传统的植被指数,这些新提出的植被指数对重金属胁迫更加敏感。虽然重金属胁迫会使植物的反射光谱发生变化,但是这种变化是十分微小的,再加上植物光谱往往还受到背景、大气、噪声的影响,导致从植物光谱中提取可以表征植物重金属胁迫的信息十分困难。近些年,Liu^[7]和 Wang^[8]分别表明对植物光谱进行小波变化可以减小背景等干扰信息的影响,增强植物受重金属胁迫的信号,并提出了一系列基于小波变换的植被指数。

(2) 植物的“三边参数”:王慧等研究表明,随着铜和锌含量的增加,小麦的红边斜率和红边峰值显著降低,相比于其他光谱参数,其对金属胁迫更加敏感^[9];朱叶青等研究土壤中铜对小麦不同生长期的光谱的影响,发现在小麦苗期和拔节苗期,随着铜的胁迫,红边会向长波方向移动,即“红移现象”,而在抽穗期和灌浆期红边会向短波方向移动,即“蓝移现象”^[10]。

(3) 某些范围下的植物叶冠光谱反射率或者吸收深度:Chi 等^[11]研究发现,随着植物叶片中的 Cu, Pb 和 Zn 的增加,植物叶绿素含量相应的减小,可见光波段反射率增加,近红外波段反射率减小。Asmeryan 等^[12]指出利用 WorldView-2 数据的第 6 波段可以较好的识别城市中的植物由于环境污染所受到重金属的胁迫程度。

在喀拉通克和希勒库都克矿区白茎绢蒿是广泛分布的一种植物。近几年已经有学者研究表明:生长在矿床上部的白茎绢蒿会吸收土壤中的金属元素在其体内形成植物地球化学异常,根据这些异常可以区分不同的矿化类型和矿化规模^[13]。虽然前人在利用植物的光谱信息发现金属胁迫方面做了大量的研究,并取得了显著地成果。但是由于不同的植物之间存在着冠层结构、内部结构等差异,导致在一种植物上总结的规律往往无法应用于其他植物,即不同植物之间存在着“尺度效应”。目前尚没有人研究是否可以根据白茎绢蒿的光谱来发现其体内的地球化学异常?如果可以?则可以把建立的模型或者总结的规律应用于航空或者航天遥感中。如

我们课题组基于动力三角翼和 HySpex 高光谱传感器构建的“超低空探测平台”(图 1),实现快速、大面积地提取研究区域植物地球化学异常。因此,本研究首次尝试从该种植物的光谱信息中挖掘出与地球化学异常密切相关的特征波段或者特征值。思路是首先通过在已知矿床上部划定采样剖面,然后对比生长在矿床上部的植物与背景区植物之间的光谱,找出它们之间的差异以及变化规律。



图 1 动力三角翼搭载 HySpex 高光谱传感器
Fig. 1 Dynamic triangular wing mounted HySpex hyperspectral sensor

1 实验部分

1.1 研究区概况

1.1.1 矿区地质背景

希勒库都克矿区位于新疆阿勒泰富蕴县境内,距离富蕴县城 40 km,是一个以钼为主的铜钼矿。目前在区域已经发现了 10 余条矿体。矿区的地质主要是石炭统南明水组凝灰质砂岩为主,夹粉砂岩、石英粉砂岩、硅质岩、灰岩。该区域断裂构造发育,以 EW, NNW 和 NE 向为主,其中 NNW 向断裂为矿区主要控矿断裂^[14]。

喀拉通克距富蕴县城 35 km,是一所大型的铜镍矿。近年来,学者们在该区域开展外围找矿和深部找矿,取得重大发现和重大突破。地层主要为下石炭统南明水组凝灰质粉砂岩、含炭质凝灰质板岩、硅质岩、中粗-粉屑凝灰岩、沉凝灰岩、含炭质凝灰岩等^[15]。该区域断裂以 NW 和 NNW 方向为主,其中 NW 方向为该区域的控岩断裂。

1.1.2 矿区植物类型

白茎绢蒿是这两个矿区广泛分布的一种植物,是多年生草本、小半灌木、高约 20~40 cm。白茎绢蒿是典型超旱生沙生小半灌木,具有发达的根系和细小的叶片,降低了水分的散失,有极强的抗旱和抗高温能力,可以生长在环境极其恶劣的区域。

1.2 采样剖线的设计

在已知矿床上部设计了 2 条剖面。喀拉通克采样剖面全长 1 700 m,包括 63 个采样点。0—30 点采样间隔为 20 m;30—39 点采样间隔为 50 m;40—44 点采样间隔为 10 m;45—55 点采样间隔为 20 m;55—62 点采样间隔为 50 m。希勒库都克采样剖面全长 1 170 m,包括 46 个采样点。0—30 点采样

间隔 20 m；30—39 采样间隔 50 m；40—45 采样间隔 20 m。包含的区域为矿床上部区。
 两条剖线的采样位置以及采样编号如图 3 所示，图中红框所

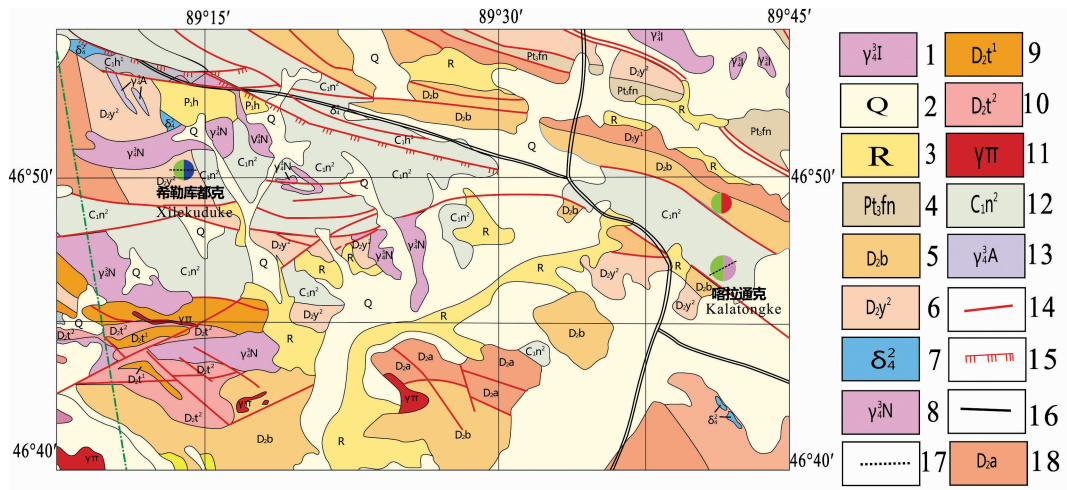
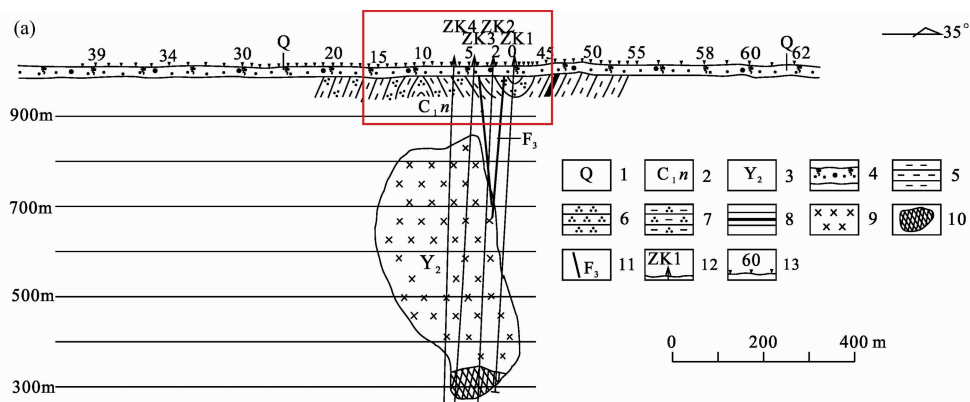


图 2 研究区域地质图

1: 海西晚期: 斜长花岗岩、混合花岗岩; 2: 第四系; 3: 泥质粉砂岩、砂岩; 4: 上元古界富蕴群: 片岩、片麻岩夹变粒岩; 5: 中泥盆统北塔山组: 中基性喷发岩、凝灰岩; 6: 中泥盆统蕴都喀拉上亚组: 砂岩凝灰岩夹酸性火山岩; 7: 海西晚期: 闪长岩类; 8: 海西晚期: 斜长花岗岩、中细粒黑云母花岗岩; 9: 下泥盆统托让格库都克下亚组: 细砂岩、粉砂岩以及粗砂岩; 10: 下泥盆统托让格库都克上亚组: 凝灰砂岩、安山玢岩; 11: 花岗岩斑岩; 12: 下石炭统南明水组: 凝灰质砂岩、凝灰岩夹酸性熔岩; 13: 海西晚期: 钾质花岗岩、云黑母花岗岩; 14: 断裂; 15: 压性、压剪性断裂; 16: 实测或者推测地界性; 17: 采样剖面; 18: 下石炭统黑山头下亚组: 安山玢岩、凝灰岩以及砂岩

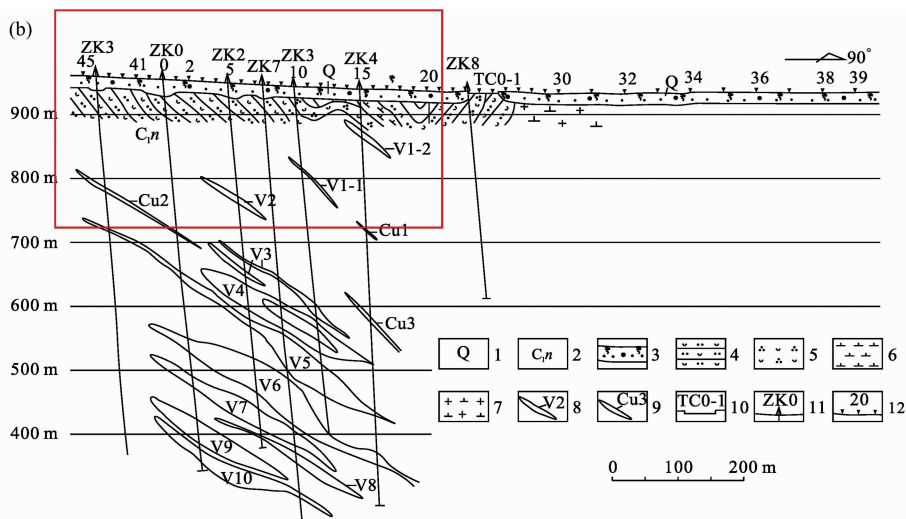
Fig. 2 The geological map of the study area

1: Late Hercynian; plagioclase granite and migmatitic granite; 2: Quaternary; 3: Argillaceous siltstone and sandstone; 4: Upper Proterozoic Schist Fuyun Formation: schist, gneiss and granulite; 5: Middle Devonian the North Tashan Formation; the intermediate basic eruptive rock and tuff; 6: Middle Devonian plagiogranite, medium fine-grained biotite granite; 9: Lower Devonian Tuoranggekuduke Lower Formation: fine sandstone, siltstone and coarse grain sandstone; 10: Lower Devonian Tuoranggekuduke Upper Formation: condensate sandstone and andesitic porphyrite; 11: Granite porphyry; 12: Lower Carboniferous Nantuo Mingshui Formation; tuffaceous sandstone, tuff and Intermediate acid lava; 13: Late Hercynian: potassic granite and biotite granite; 14: Fault; 15: Compressional and compressive shear fracture; 16: Measured or conjectural boundaries; 17: Sampling lines; 18: Lower Devonian Heishantou Lower Formation: andesitic porphyrite, tuff and sandstone



1: 第四系; 2: 下石炭统南明水组; 3: 杂岩体编号; 4: 砂砾质土; 5: 沉积岩; 6: 中粗屑凝灰岩; 7: 中粗粉屑凝灰岩; 8: 凝灰质斑岩; 9: 基性杂岩; 10: 铜镍矿体; 11: 断裂; 12: 钻孔位置; 13: 采样位置以及编号

1: Quaternary; 2: Lower Carboniferous Nantuo Mingshui Formation; 3: The number of complex rock mass; 4: Gritty soil; 5: Sedimentary tuff; 6: Middle coarsely tuff; 7: Medium coarse powder tuff; 8: Tuffaceous slate; 9: Basic complex; 10: Copper-nickel ore body; 11: Fault; 12: The position of the drills; 13: The location and number of sampling points



1: 第四系; 2: 下石炭统南明水组; 3: 砂砾质土; 4: 凝灰质砂岩; 5: 石英斑岩; 6: 闪长岩; 7: 花岗闪长岩; 8: 钼矿体; 9: 铜矿体; 10: 探槽; 11: 钻孔; 12: 采样位置以及编号

1: Quaternary; 2: Lower Carboniferous Nantuo Mingshui Formation; 3: Gritty soil; 4: Tuffaceous sandstone; 5: Quartz porphyry; 6: Diorite; 7: Granodiorite; 8: Molybdenum ore bodies; 9: Copper ore bodies; 10: Trench; 11: Drills; 12: The location and number of sampling points

图 3 采样剖线的设计以及采样点的位置

(a): 喀拉通克矿区; (b): 希勒库都克矿区

Fig. 3 Sampling lines and location of sampling points

(a): Kalatongke mining area; (b): Xilekuduke mining area

1.3 植物光谱的采集

ASD Fieldspec3 是一种便携式的地物光谱仪, 使用该光谱仪可以快速地获取植物在 350~2 500 nm(可见光、近红外与中红外)范围内的反射率。为了减少外部因素的干扰, 获得稳定的数据, 应该选择晴朗、无风的天气, 并且在当地时间的 10:30—14:00 进行测量, 因为在此时间段内太阳高度角比较稳定。

在本次实验中裸探头的视场角为 25°, 考虑到白茎绢蒿的冠幅的大小, 因此探头距离冠层顶部在 20 cm 比较合适。为了保证在同一个样品光谱采集的过程中减少人为以及自然环境的干扰, 每一个样品采集 5 条光谱, 然后剔除变异比较大的光谱, 剩余的光谱求平均数作为该样品的反射光谱。

1.4 植物反射光谱数据的预处理

由于小于 400 nm 和大于 2 400 nm 范围波段的信噪比较低, 受到外界干扰光谱数据的波动性较大, 因此将这两个范围内的数据删除。1 300~1 400 和 1 800~2 000 nm 范围内的数据受大气水汽的影响较大, 反射率往往出现异常(>1), 因此这两个波段的数据同样被删除。最终剩余的波段数为 1 700 个, 波段宽度为 1 nm。但是过窄的波段宽度会造成数据的冗余, 同时植物光谱会产生“毛刺”现象, 因此本文通过求相邻的 10 个波段的平均值来对植物光谱进行平滑处理。通过平滑处理不仅去除了一部分噪声的影响, 同时也最大程度保留了植物光谱的波形, 最终植物的光谱剩下 170 个波段。处理以后的植物反射光谱如图 4 所示。

1.5 一阶导数和二阶导数

相关研究表明微分光谱可以有效的减小土壤背景等干扰因素对冠层光谱的影响, 反映植物内部物质的吸收的波形变

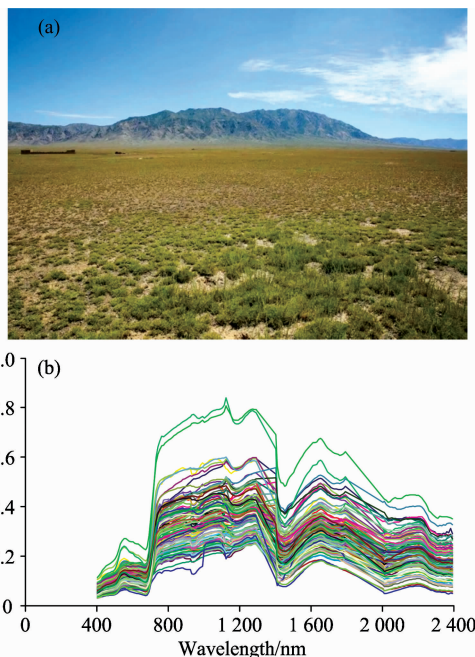


图 4 矿区的自然景观以及采集的植物样品的光谱曲线
Fig. 4 Natural landscape and reflectance spectral curve of plant samples collected in the mining area and background area

化, 从而更有利于从植物冠层光谱中提取生化信息。相比于原始光谱, 导数光谱可以探测到植物光谱中的一些微小变化, 起到“放大镜”的作用。因此, 不仅比较背景区和矿区的

植物的原始光谱,同时比较其一阶导数和二阶导数,找出差异显著的波段或者特征值。由于植物光谱数据实质是“离散型”数据,因此导数变换使用差分方法近似计算,一阶求导和二阶求导的计算公式分别如式(1)和式(2)所示

$$R'(\lambda_i) = \frac{R(\lambda_{i+1}) - R(\lambda_{i-1})}{2\Delta\lambda} \quad (1)$$

$$R''(\lambda_i) = \frac{R'(\lambda_{i+1}) - R'(\lambda_{i-1})}{2\Delta\lambda} \quad (2)$$

式中, $R(\lambda_i)$ 为 λ_i 处的波段反射率; $R'(\lambda_i)$ 为 λ_i 处的一阶导数值; $R''(\lambda_i)$ 为 λ_i 处的二阶导数值; $\Delta\lambda$ 为波长间隔。

1.6 分形维数

分形维数一般被用来描述自然界中复杂的图形的整体特征。目前,一些研究表明地物的光谱曲线也具有分形特征,同时过往的研究提出的一些光谱特征参数无论是植物指数还是“三边”参数都只针某一段光谱曲线,而往往忽略了植物光谱曲线的整体特征,导致丢失了很多重要的信息。因此引入分形理论来综合描述生长在矿床上部的植物和背景区的植物的光谱差异,并将光谱曲线的分形维数作为一种新的光谱特征参数。对于分形维数的计算,使用盒维法,其计算公式如式(3)所示

$$D = -\lim_{r \rightarrow 0} \frac{\ln N_r}{\ln r} \quad (3)$$

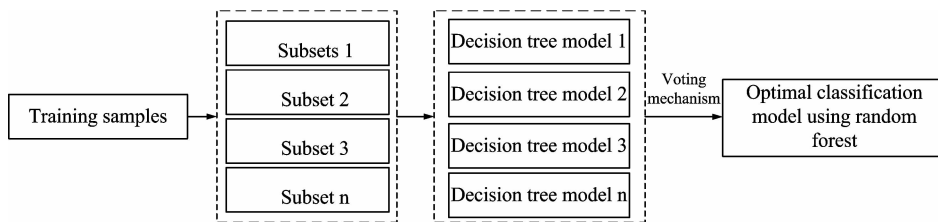


图 5 随机森林分类原理

Fig. 5 The principle of random forest classification

1.9 技术路线图

技术路线图如图 6 所示,研究步骤主要包括:(1)在喀拉通克和希勒库都克的已知矿床上部划定两条剖面。

(2)沿着剖面进行植物反射光谱的采集。在希勒库都克矿区,矿床上部的采样间隔为 20 m,背景区的采样间隔为 50 m;在喀拉通克矿区,矿床上部的采样间隔为 10~20 m,在背景区的采样间隔为 50 m。最终在希勒库都克矿区采集了 45 个植物样品,其中 0~20, 40~45 在矿床上部,其余在背景区;在喀拉通克矿区采集了 62 个植物样品,其中 0~16, 40~45 在矿床上部,其余在背景区域(如图 3 所示)。

(3)对生长在矿床上部的植物和背景区的植物分别进行一阶求导、二阶求导以及分形维数的计算;然后对比分析找出差异显著的波段。

(4)首先把矿床上部的区域记为 1,背景区域记为 0,然后把优选出特征波段分别作为 PLS-SVM 模型和随机森林模型的输入参数建立分类模型,并对这两种分类模型去一交叉验证分类精度进行比较找出最优模型,最终形成一套快速地利用植物光谱寻找隐伏矿床的方法。

式中: r 为正方形网格的边长;用边长为 r 的正方形网格覆盖光谱曲线, N_r 为光谱曲线与边长为 r 的正方形网格相交的非空网格数;逐步改变正方形网格的大小会得到一系列 N_r 值; D 则是分形维数。

1.7 偏最小二乘-支持向量机分类(PLS-SVM)

偏最小二乘是一种新型的数学统计方法,它是将变量之间的相关分析和多元线性回归结合在一起,可以在自变量严重多重相关的条件下进行建模,同时也可以在本样本个数小于变量个数的情况下进行建模,因此偏最小二乘在地学领域得到了越来越广泛地应用。

支持向量机在解决小样本、非线性方面表现出许多特有的优势,已经被广泛地应用于回归和分类中。偏最小二乘-支持向量机可以充分发挥最小二乘和支持向量机的优势。

1.8 随机森林分类

随机森林是通过构建多棵决策树的分类结果进行投票来决定新样本的类别,克服了使用单一准则进行分类的偶然性以及不稳定性。

随机森林的原理是使用 Bootstrap 方法从训练数据集中有放回的抽取多个样本组成子数据集,然后每个子数据集进行决策树建模,然后对多棵决策树的分类结果进行投票确定最优的分类结果,其原理如图 5 所示。

2 结果与讨论

2.1 差异显著的波段的选择

为了从众多的波段中选择出差异显著的波段,以 T 检验的 p 值作为指标对波段进行筛选。但是 T 检验属于参数检验,参数检验要求样本服从正态分布且这些正态总体拥有相同的方差。因此在做 t 检验之前,对样本要先进行正态性检验和方差齐性检验,这些步骤均可以在 matlab 内实现,最终以 p 是否小于 0.01 作为该波段是否是差异显著的波段。从表 1 中我们可以看出对于原始光谱,矿床上的植物光谱和背景区相比不存在差异显著的波段;而进行一阶求导以后差异显著的波段数目增加到了 6 个,分别为 824, 834, 1 533, 1 573, 1 633 和 1 643,进行二阶求导以后差异显著的波段数量较少,只有两个分别为 1 284 和 1 703,一阶导数和二阶导数的分形维数甚至通过了 $p=0.001$ 的检验,这些特征波段可以最大程度上发现生长在矿上的植物和背景区的植物的差异,作为植物地区寻找隐伏矿床的植物地球化学标志。

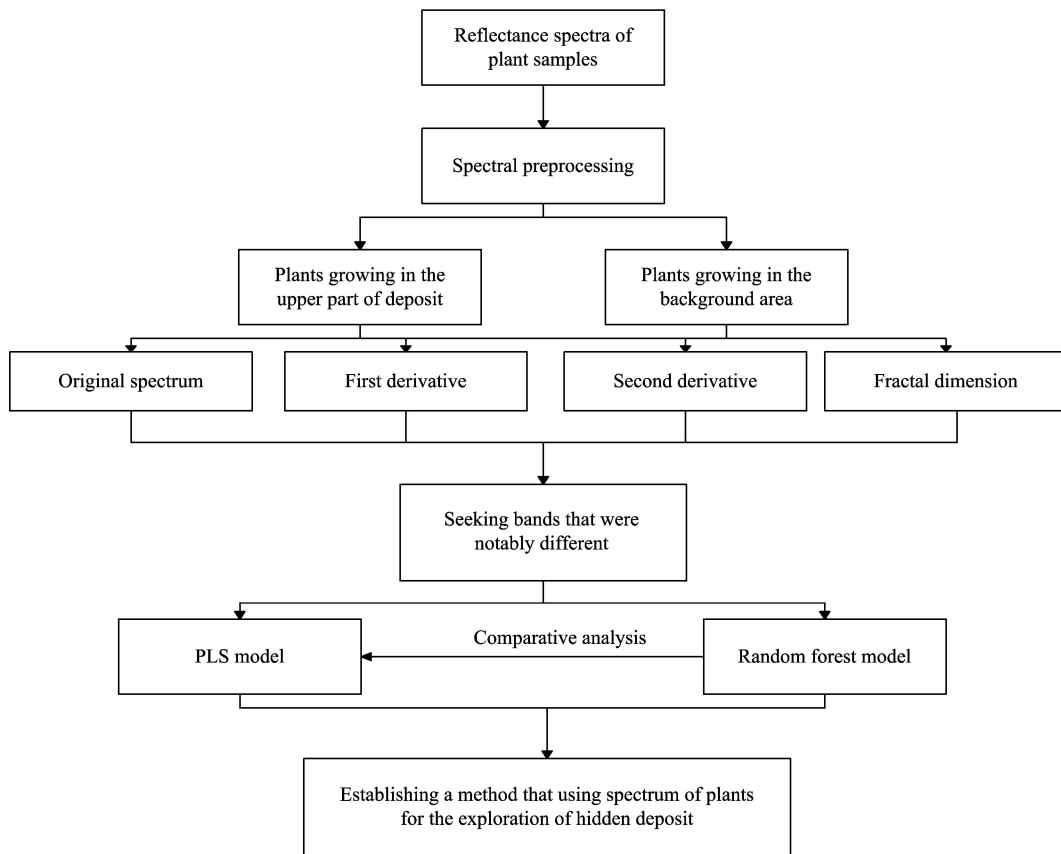


图 6 技术路线图

Fig. 6 Technology roadmap

对于植物光谱，可见光范围(400~700 nm)主要受到色素的影响；近红外波段(700~1 300 nm)主要受到细胞结构的影响；中红外波段(1 300~2 400 nm)主要受到含水量的影响。已有研究表明^[16-17]：生长在金属矿床上部的植物的叶细胞中会聚集大量的电子致密性物质，使得植物的细胞内部结构遭受一定的损坏，这可能就是矿床上生长的植物的光谱相比于背景区差异显著的波段位于近红外波段范围的原因。

表 1 表明：差异显著的波段集中于近红外和中红外波段，而在可见光范围，矿床上生长的植物的光谱相比于背景区没有显著性差异。通过实地调查，我们也发现生长在矿床上的植物的颜色相比于背景区没有显著的差异。相关研究认为：生长在矿区的植物呈现黄绿色，而生长在背景区的植物呈现青绿色，根据色彩的异常圈定成矿靶区^[18]。我们的研究结果与他们的有一定的差异，造成这种现象的原因可能是不同的植物对重金属的吸收有不同的“屏障效应”，本研究对象(白茎绢蒿)对重金属的忍耐程度大于前人的研究对象，虽然生长在矿床上吸收了重金属元素，但是还没有出现中毒症状。

搭载于航空和航天平台上的高光谱传感器对地物是垂直观测。为了可以将本研究建立的模型更好地应用于航空和航天平台上。我们也使用 ASD 光谱仪对植物进行垂直观测，获取植物的冠层光谱。但是由于白茎绢蒿是一种小半灌木，叶片较小，叶面冠层的覆盖度无法达到100%，导致获取的植

表 1 不同的光谱变换形式下通过 $p < 0.01$ 的波段

Table 1 The band at $p < 0.01$ test level by different spectral transformations

	波段位置	p 值
一阶导数	R'_{824}	0.009 013
	R'_{834}	0.007 844
	$R'_{1\ 573}$	0.009 163
	$R'_{1\ 633}$	0.005 234
	$R'_{1\ 643}$	0.005 88
	$R'_{1\ 533}$	0.007 09
二阶导数	$R''_{1\ 284}$	0.008 979
	$R''_{1\ 703}$	0.006 989
分形维数	一阶导数	0.000 000 680 6
	二阶导数	0.000 009 76

物光谱中包含土壤信息。也就是说获取的植物光谱其实是一个混合光谱，既包含了植物的光谱信息，同时也包含了土壤信息。当使用植物的原始光谱进行对比分析的时候，土壤的影响较大。而土壤光谱在可见光、近红外以及中红外范围反射率变化不大，近似于一条直线。因此土壤光谱的加入往往会降低植物光谱之间的差异。也就是为什么生长在矿床上部的植物的原始光谱与背景区相比，不存在差异显著的波段。而经过一阶以及二阶导数以后，土壤光谱被消除，植物光谱被保留，这样相比于原始光谱就扩大了不同区域之间的光谱

差异。同时由于导数光谱可以有效地消减土壤背景、大气水汽含量等干扰因素的影响,使得基于导数光谱建立的预测模型具有更好的鲁棒性。

光谱变换是一种有效的消除噪声,提取特征信息的方法。目前比较常用的光谱变换方法包括,导数光谱、对数光谱、倒数光谱等。在未来的研究中我们可以尝试在植物的可见光、近红外以及中红外范围内分别寻找最优的光谱变换方式,然后在整个植物光谱范围形成一个组合光谱,最大程度上凸显光谱之间的差异。

2.2 差异显著的特征波段的变化趋势

2.1 节中优选出了矿区植物光谱相比于背景区差异显著的特征波段,本节主要是比较生长在不同区域的植物的这些特征波段的变化趋势。图 7 和表 2 显示出无论是一阶导数、二阶导数还是分形维数,矿床上部相比于背景区域都有一定的增加,其中二阶导数的增长幅度最大,一阶导数的增长幅度也在 17.8%~22.4%之间,而分形维数的增长幅度最小。这 10 种差异显著的特征波段的变化趋势相同,因此可能可以使用这 10 种特征波段组合成一个“复合指数”作为寻找隐伏矿床的植物地球化学标志,但是采用何种方法对这 10 种特征波段进行合成仍然需要进一步研究。衬度系数是指矿区中植物的某参数与背景区的比值。它越大,表明越可以清晰

地凸显出异常,对异常信息的挖掘越有利。从表 2 可以看出,二阶导数的衬度系数明显大于一阶导数,这说明二阶导数可以更好地突出生长矿区中的植物异常信息,是更好的植物地球化学找矿标志。因此,相比于一阶求导,对植物的光谱进行二阶求导是更好的光谱变换方式。

表 2 差异显著的特征波段的变化幅度

Table 2 Variation range of significant characteristic bands

	特征波段	背景区	矿床上部	增长幅度 /%	衬度系数
一阶导数	R'_{824}	0.000 268	0.000 315	17.818 96	1.18
	R'_{834}	0.000 276	0.000 334	20.755 53	1.21
	R'_{1573}	0.000 2	0.000 233	16.806 56	1.17
	R'_{1633}	0.000 158	0.000 188	18.925 58	1.19
	R'_{1643}	0.000 118	0.000 144	22.358 79	1.22
	R'_{1533}	0.000 264	0.000 316	19.824 18	1.20
二阶导数	R''_{1284}	-0.000 016	-0.000 026	62.5	1.63
	R''_{1703}	0.000 001 17	0.000 002 5	114	2.14
分形维数	一阶导数	1.232 2	1.266 55	2.79	1.03
	二阶导数	1.217 916	1.249 23	2.57	1.03

生长在矿床上部的植物与背景区的植物相比,一阶导数的分形维数差异量略大于二阶导数,增强了分形维数提取光谱弱信息变化的能力,因此相比于二阶导数,一阶导数的分形维数更适合于作为植物地球化学找矿标志。

2.3 预测模型的建立

在使用 PLS-SVM 进行建模的过程中,选择径向基函数作为支持向量机的核函数,最优正则化参数(γ)和核参数(σ^2)使用网格搜索快速留 1 交叉验证确定。把 2.1 节优选出的 10 种差异显著的特征波段进行归一化处理以后作为输入参数代入到 PLS-SVM 建立模型。结果显示:拟合预测分类准确率达到 93%,去一预测分类准确率也达到了 84%。

决策树个数是影响随机森林分类精度的重要因素,个数过多会使得模型过度拟合,降低模型的鲁棒性。因此使用留一交叉验证的方法,确定最优决策树个数为 21~25。从表 4 中可以看出,使用随机森林的方法虽然拟合预测精度相比于 PLS-SVM 有了提高,但是去一交叉验证的分类精度从 84% 下降到了 78%。表明相比于随机森林,利用 PLS~SVM 建立的模型泛化能力更强、预测能力更强。

(1)使用 PLS-SVM 和植物光谱进行隐伏矿床的探查虽然取得了较好的效果,但是岩石和土壤中地球化学信息对隐伏矿床也同样具有指示性作用,因此可以尝试建立一个基于植物、岩石以及土壤多源信息相结合的预测模型。

(2)由于不同数据源之间存在空间和光谱尺度差异,因此把地面建立的模型应用于航空和航天数据之前,需要做尺度转化。但是如何进行不同数据之间的尺度转换需要进一步研究。

(3)基于高光谱数据的植物异常信息提取是否存在“时间尺度”需要进一步研究。也就是说对于白茎绢蒿,受重金属胁迫的与不受重金属胁迫的植物相比,什么季节它们之间

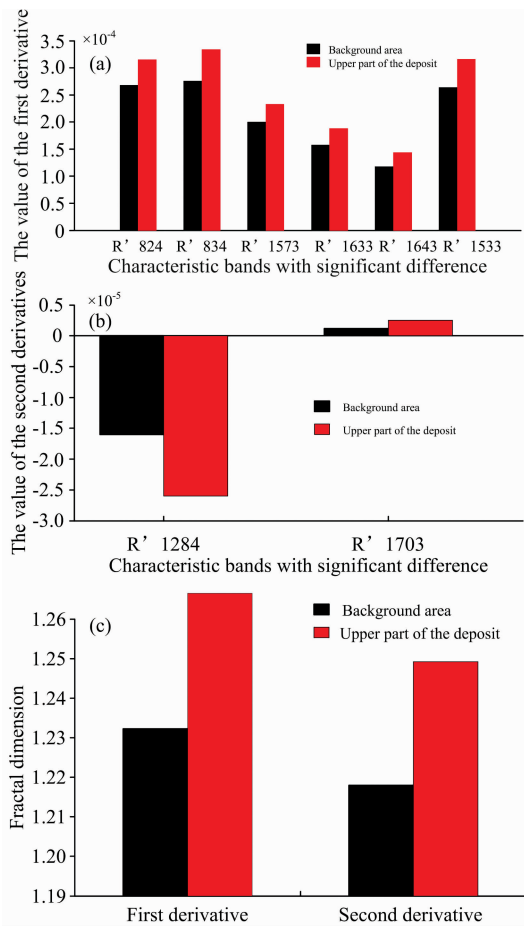


图 7 差异显著的特征波段对比

Fig. 7 Comparison of characteristic bands with significant difference

表 3 不同建模方法精度比较

Table 3 The accuracy comparison of different modeling methods

模型类型	最优正则化参数	最优核参数: σ^2	核函数	拟合预测 分类精度/%	留一交叉验证 分类精度/%
PLS-SVM	11.42	78.74	RBF	93	84
随机森林		最优决策树个数 21		97	78

差异更明显,可以最大程度上的凸显出异常信息,尚待深究。

3 结 论

从原始光谱、一阶导数光谱、二阶导数光谱以及分形维数四个角度对比了生长在矿床上的植物和背景区植物光谱的差异,然后优选出差异显著的波段或者因子,最后分别使用随机森林和偏最小二乘-支持向量机建立基于植物光谱的隐伏矿床预测模型,得到如下结论:

(1)生长在矿上的植物与背景区相比,优选出了 10 个差异显著波段/特征值,分别为: R'_{824} , R'_{834} , R'_{1533} , R'_{1573} ,

R'_{1633} , R'_{1643} , R''_{1284} , R''_{1703} , 一阶导数的分形维数以及二阶导数的分形维数,这些特征波段可以作为植物地区寻找隐伏矿床的植物地球化学标志。

(2)与原始光谱相比,进行一阶求导和二阶求导以后可以增大矿床上植物和背景区植物之间的差异。

(3)生长在矿床上植物与背景区相比,优选出 10 个差异显著的波段具有明显的增大趋势。

(4)以优选出的 10 个差异显著的特征波段作为输入参数,使用随机森林和偏最小二乘-支持向量机建立隐伏矿床预测模型均可以取得较好的效果。相比于随机森林,偏最小二乘-支持向量机建立的模型具有更好的泛化能力。

References

- [1] Rathod P H, Brackhage C, Meer F D V D, et al. European Journal of Remote Sensing, 2015, 48(3): 283.
- [2] Sridhar B B, Vincent R K, Roberts S J, et al. International Journal of Applied Earth Observation & Geoinformation, 2011, 13(4): 676.
- [3] Hede A N H, Kashiwaya K, Koike Katsuki, et al. Remote Sensing of Environment, 2015, 171: 83.
- [4] Zhang C, Ren H, Qin Q, et al. Remote Sensing Letters, 2017, 8(6): 576.
- [5] Zhang B, Wu D, Zhang L, et al. Environmental Earth Sciences, 2012, 65(3): 649.
- [6] Shi T, Liu H, Chen Y, et al. Journal of Hazardous Materials, 2016, 308: 243.
- [7] Liu M L, Liu X N, Wu M X, et al. Computers & Geosciences, 2011, 37: 1642.
- [8] Wang J J, Wang T J, Shi T Z, et al. Remote Sensing, 2015, 7: 15340.
- [9] WANG Hui, ZENG Lu-sheng, SUN Yong-hong, et al(王 慧, 曾路生, 孙永红, 等). Transactions of the Chinese Society of Agriculture Engineering(农业工程学报), 2017, 33(2): 171.
- [10] ZHU Ye-qing, QU Yong-hua, LIU Su-hong, et al(朱叶青, 屈永华, 刘素红, 等). Journal of Remote Sensing(遥感学报), 2014, 18(2): 335.
- [11] Chi G Y, Shi Y, Chen X, et al. Advanced Materials Research, 2012, 347: 2735.
- [12] Asmaryan S, Warner T A, Muradyan V, et al. Remote Sensing Letters, 2013, 4(2): 200.
- [13] SONG Ci-an, SONG Wei, DING Ru-fu, et al(宋慈安, 宋 玮, 丁汝福, 等). Geotectonica et Metallongenica(大地构造与成矿学), 2017, 41(1): 122.
- [14] YOU Jun, HONG Tao, WU Chu, et al(游 军, 洪 涛, 吴 楚, 等). Acta Petrologica Sinica(岩石学报), 2016, 32(5): 1262.
- [15] QIN Ke-zhang, TIAN Ye, YAO Zhuo-sen, et al(秦克章, 田 野, 姚卓森, 等). Geology in China(中国地质), 2014, 41(3): 912.
- [16] Ge W, Jiao Y Q, Sun B L, et al. South African Journal of Botany, 2012, 83(4): 98.
- [17] Caldelas C, Bort J, Febrero A. Cell Biology and Toxicology, 2012, 28(1): 57.
- [18] Filippidis A, Papastergios G, Kantiranis N et al. Chemie der Erde-Geochemistry, 2012, 72(1): 71.

Extraction of Plant Abnormal Information in Mining Area Based on Hyperspectral

CUI Shi-chao^{1, 2, 3}, ZHOU Ke-fa^{1, 2*}, DING Ru-fu⁴

1. Xinjiang Research Center for Mineral Resources, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi 830011, China
2. Xinjiang Key Laboratory of Mineral Resources and Digital Geology, Urumqi 830011, China
3. University of Chinese Academy of Sciences, Beijing 100049, China
4. China Non-Ferrous Metals Resource Geological Survey, Beijing 100012, China

Abstract *Seriphidium terrae-albae* is a kind of plant widely distributed in various mining areas of Fuyun County, Xinjiang, China. The traditional exploration methods are difficult to play a role due to the existence of plant information and other obstacles, and some new methods and new ideas are urgently needed. The remote sensing plant geochemistry method is a kind of natural information source that smartly utilizes plants, transforming the plant from the barrier information to the useful information. Help people quickly and economically obtain the useful information about minerals under plant barriers. Because of its large area, being fast and non-destructive and other advantages, it has attracted more and more attention of scholars, and has become the current research hotspot. In recent years, although some scholars have synthetically considered “absorption coefficient” and “contrast coefficient” to prove that *Seriphidium terrae-albae* can be used as a good indicator for the exploration of concealed deposits. The plants in the upper part of the deposit can absorb the ore-forming elements in the soil better, but at the same time they form geochemical anomalies in their bodies, and the information is more visible than other plant anomalies. However, no one has studied whether the geochemical anomalies in *Seriphidium terrae-albae* can be found from the spectral point of view, then providing some references for the exploration of concealed deposits. Therefore, our study first tries to look for the feature bands or eigenvalues closely related to geochemical anomalies, and then construct the prediction model of hidden deposit based on plant spectrum. First, the method adopted was to measure the reflectance spectra of plants growing in the upper part of deposit and background area by ASD FieldSpec3 spectrometer respectively. Then the spectra of plants growing in these two regions were analyzed and compared from five aspects, including the original spectrum, the first derivative spectrum, the second derivative spectrometry, the first derivative fractal dimension and the second derivative fractal dimension. Finally the 10 characteristic bands that were notably different were selected including R'_{824} , R'_{834} , R'_{1533} , R'_{1573} , R'_{1633} , R'_{1643} , R''_{1284} , R''_{1703} , the first derivative fractal dimension and the second derivative fractal dimension. These characteristic bands can be used as botanogeochemistry marks for seeking exploration of concealed deposits. Taking these ten optimized bands as input parameters, random forest (RF) and partial least squares support vector machine (PLS-SVM) were used to construct a prediction model that seeks the position of hidden deposits based on abnormal spectrum of plant. The results showed that these two models can obtain satisfactory results, but compared with the random forest model, the partial least squares support vector machine model has a better robustness and stronger generalization ability. The results also indicated that it has great potential in looking for hidden deposit using extraordinary spectrum of plants, due to the advantages of being simple and quick. Our team has built a “very low altitude detection platform” using dynamic delta wing and HySpex imaging hyperspectral sensor, which can realize the observation of “sub-meter”. But the problems will be our next research focus as follows, how to effectively solve the problem of “spatial scale” and “spectral scale”? How to better apply the model established on the ground test ground to the very low altitude detection platform, and how to extract the plant anomaly information in a large area and quickly in the research area?

Keywords Hyperspectral; Hidden deposits; Random forest (RF); Partial least squares support vector machine (PLS-SVM); Botanogeochemistry marks

(Received Dec. 11, 2017; accepted May 5, 2018)

* Corresponding author