引文格式：刘翰霖,李根军,林楠,等. 基于极端随机树的高光谱蚀变信息提取研究[J]. 地理空间信息, 2023, 21(9): 57-60. doi:10. 3969/j.issn. 1672-4623.2023.09.014

基于极端随机树的高光谱蚀变信息提取研究

刘翰霖[1]，李根军[2,3]，林　楠[1]，杨雪松[2]
（1. 吉林建筑大学，吉林 长春 130118；2. 青海省地质调查院，青海 西宁 810008；
3. 青海坤拓遥感技术服务有限公司，青海 西宁 810007）

摘　要：以青海沟里地区为研究区，对区内资源一号02D（ZY1-02D）高光谱遥感影像进行预处理，充分分析了区内地质矿产资料，利用光谱角匹配技术提取典型蚀变矿物光谱样本数据，利用极端随机树算法开展矿化蚀变信息提取研究，并对提取效果进行评价。结果表明，提取的蚀变矿物信息分布位置与区内断裂构造展布特征以及已知矿床（点）的分布规律基本吻合，符合研究区成矿地质特征，说明结合光谱角匹配技术和极端随机树算法的蚀变矿物提取模型是一种高效的矿产勘查方法。

关键词：高光谱遥感；ZY1-02D数据；蚀变矿物；极端随机树
中图分类号：P237　　**文献标志码：**B　　**文章编号：**1672-4623（2023）09-0057-04

Hyperspectral Alteration Information Extraction Based on Extremely Randomized Tree

LIU Hanlin[1], LI Genjun[2,3], LIN Nan[1], YANG Xuesong[2]
(1. Jilin Jianzhu University, Changchun 130118, China;
2. Qinghai Geological Survey Institute, Xining 810008, China;
3. Qinghai Kuntuo Remote Sensing Technology Service Co., Ltd., Xining 810007, China)

Abstract: Taking Gouli area of Qinghai Province as the study area, we preprocessed the Ziyuan-1-02D (ZY1-02D) data hyperspectral remote sensing image in the area. Based on the analysis of geological and mineral data in the area, we extracted the spectral sample data of typical altered minerals based on spectral angle matching technology, used the extremely randomized tree algorithm to extract mineralization alteration information, and evaluated the extraction effect. The result shows that the distribution position of altered mineral information extracted by the model is basically consistent with the distribution characteristics of fault structures in the area and the distribution law of known deposits(points), which is in line with the metallogenic geological characteristics of study area, indicating this model is an efficient mineral exploration method.

Key words: hyperspectral remote sensing, ZY1-02D data, altered mineral, extremely randomized tree

随着矿产资源的高强度急速消耗，多处矿山因长期透支开采逐渐枯竭，利用传统的地球物理、地球化学方法进行矿产资源勘查，不仅费时费力、经济成本巨大，而且经常会出现生态环境污染问题。因此，在矿产资源勘查和生态环境保护双重目标下，如何利用高新技术准确、快速、有效地开展绿色勘查工作，减少地表工程揭露，保护周边生态环境，是目前矿产资源勘查急需解决的首要问题[1]。高光谱遥感技术能在准确区分岩矿类型的同时，定量分析矿物成分，是一种快速、大面积、低成本观测岩矿物质组成的有效技术手段[2-4]。随着统计学习理论在遥感地质领域研究的不断深入，支持向量机、随机森林、多层感知神经网络等机器学习算法逐渐被引入岩性信息识别和蚀变矿物提取中，并取得了一定的研究成果；但上述方法针对高光谱遥感多源数据，通常存在计算复杂度较高、模型过拟合等问题，因此有必要探索一种新方法来实现高光谱蚀变信息的高精度分离。极端随机树是在随机森林基础上发展而来的新型机器学习算法，具有较高的鲁棒性和泛化能力，能在保证算法稳定性的同时加快复杂数据集合的运算速率，精度更高[5-7]。目前，极端随机树算法的优越性能已引起科学研究领域的广泛重视，被应用于医学、经济学、计算机科学等领域，但在遥感蚀变信息的相关研究中还较少见[8-10]。鉴于此，本文以青海沟里地区为研究对象，基于资源一号02D（ZY1-02D）高光谱遥感影像数据，结合蚀变矿物波谱特征，利用极端随机树算法提取矿化蚀变信息，并通过已知地质矿产资料评价和分析提取效果。本文成果能促进高光谱遥感技术在矿产勘查中的有效应用，为青海沟里地区矿产资源高效开发利用和生态环境保护提供技术保障。

收稿日期：2022-05-03；修回日期：2022-07-28。
项目来源：国家自然科学基金资助项目（41702357）；青海省科技厅科技成果转化专项资助项目（2020-SF-150）。
第一作者简介：刘翰霖（1997—），硕士研究生，工程师，主要从事遥感地质应用研究工作，E-mail：952647339@163.com。

1 区域地质概况

青海沟里地区地处柴达木盆地南缘，位于东昆仑矿山活动断裂带东段，是我国重要的金矿产地，素有“金腰带”美称。该地区多旋回造山运动极为发育，曾历经多期次扩张—收缩和俯冲造山作用，使得出露地层单元呈现成层无序的组合特征，产生时代跨度大的独特分布趋势，广泛存在于太古宇至新近纪的各个地质时期。区内构造活动频繁，以压性或压扭性断裂为主，走向为北西西向，多表现为逆冲特征，包括北界断裂、狼日扎岗—龙里—尕之麻断裂和东昆中断裂等；岩浆活动强烈，侵入岩种类丰富，主要集中于加里东期—印支造山旋回的不同阶段，由奥陶纪、志留纪、二叠纪和晚三叠纪中酸性侵入岩组成，以花岗岩类侵入岩巨量出露为典型特征。研究区矿产资源丰富，以金、银、铅、锌、铜为主，其中岩浆热液型金矿床分布较广，褐铁矿、绿泥石、方解石、高岭土等蚀变矿物在区内地表出露面积较大，且光谱特征典型，可利用高光谱遥感影像开展分类提取实验研究，对进一步的矿产勘查和矿产潜力评价具有指示意义。

2 数据来源和研究方法

2.1 数据来源

ZY1–02D卫星携带有可见光近红外相机和高光谱相机，2019年10月发射升空，轨道高度为778 km，重访时间为3 d。ZY1–02D高光谱数据共包含166个波段，光谱范围为0.40～2.50 μm，光谱分辨率在可见光和短波红外分别为10 nm、20 nm，空间分辨率为30 m，可有效获取地球表面地物光谱信息，为蚀变矿物提取提供可靠的数据支持。本文选取覆盖青海沟里地区的3景ZY1–02D高光谱数据（L1A0000172721、L1A0000172720、L1A0000181506），卫星过境时间为8—10月，该时间段内地面无积雪覆盖。

2.2 数据预处理

由于原始影像上记录地物信息的灰度值是一个无量纲数值，需按照给定的卫星定标文件读入每个波段的波长范围、中心波长和信息增益值，利用ENVI 5.3软件的辐射定标模块自动提取定标参数，并将卫星搭载的成像光谱仪接收的DN值转化为大气外层表观反射率[11]。本文采用MODTRAN 5理论模型，通过ENVI软件的FLAASH模块对影像进行大气校正[12]。为保证遥感影像的几何精度、修正几何畸变，将控制点拟合中误差设定为1.5～2个像元，在每景影像上选取分布均匀的15个控制点，利用二次多项式纠正法对ZY1–02D影像进行几何校正，消除影像几何位置误差[13]。由于云雾和信息缺失等因素影响，几何校正后的影像数据并不能直接用于实验研究[14]，本文分别对影像进行几何镶嵌、灰度镶嵌，得到覆盖研究区的完整影像。

2.3 高光谱遥感蚀变信息提取

2.3.1 基于光谱角匹配技术的样本选取

精确选择训练样本是利用机器学习算法进行矿物识别的关键。为增强训练样本的代表性和可靠度，本文采用光谱角匹配技术选取训练样本。光谱角匹配是通过计算目标光谱与参考光谱间广义夹角来表征光谱相似程度的监督分类方法，通常夹角越小，相似度越大，代表匹配效果越佳[15–16]，数学表达式为：

$$\theta=\cos^{-1}[\sum_{i=1}^{n}t_i r_i/(\sqrt{\sum_{i=1}^{n}t_i^2}\times\sqrt{\sum_{i=1}^{n}r_i^2})] \quad (1)$$

式中，n为高光谱数据的波段数；t_i为目标光谱第i波段的反射率值；r_i为参考光谱第i波段的反射率值。

以USGS光谱库中的褐铁矿、绿泥石、方解石和高岭土矿物光谱为标准光谱，将光谱按照ZY1–02D影像波长范围进行重采样；基于光谱角匹配技术，以0.15弧度为分割阈值，分别提取研究区内4种典型矿物像元的光谱，以此为模型建模的光谱样本数据集，褐铁矿、绿泥石、方解石、高岭土分别提取了267个、179个、235个、187个端元光谱。

2.3.2 基于极端随机树的矿化蚀变信息提取

极端随机树算法最早由比利时列日大学电气工程与计算机科学系的Geurts P教授在2006年提出，是随机森林算法的推广形式[17]，二者都具有不易过拟合、抗噪性能良好、运算效率高等优点，但在选取训练样本和分裂节点方面存在差异。随机森林算法采用有放回抽样方法获取样本数据，通过随机组建特征属性集合，筛选最优属性作为分裂节点进行决策树训练，存在训练数据集重复性升高、样本数据利用率降低和基学习器间相似程度加深等缺陷；而极端随机树算法则利用未进行重采样的全部样本数据，选取随机生成的分类节点进行分裂，基于“误差—分歧分解”理论，依靠多颗决策树的组合来完成模型的训练过程[18]，对于小样本数据集具有更好的平滑性，在处理强噪声和高维模式的信息识别和分类问题中具有独特优势[19–20]。极端随机树算法原理见图1。

高光谱影像存在波段数量多、数据体量大、波段间信息冗余度高和共线性强等问题，为消除波段冗余信息对极端随机树算法识别效果的影响，本文基于主成分分析法，通过正交变换将共线性高的波段数据转换为线性不相关变量来提取主成分，以累积贡献率95%为阈值，基于光谱角提取的训练样本，分别对

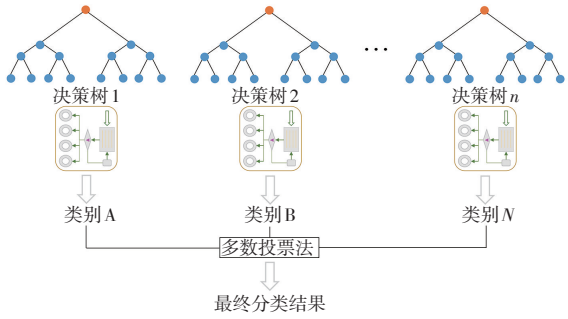


图1　极端随机树算法原理

4种蚀变矿物进行特征波段筛选，结果表明经主成分分析法降维后的波段数占总波段数的10.60%~11.92%，褐铁矿、绿泥石、方解石和高岭土的光谱波段数分别被降至16个、17个、16个和18个，降低了波段间光谱信息的冗余性和模型计算的复杂性。

准确、快速、高效地确定模型的内部关键参数，能有效提高模型分类精度和泛化性能。CART决策树是极端随机树算法的基学习器，其生长深度（树深）和数量是极端随机树算法的两个重要建模参数。本文利用网格搜索—交叉验证方法分别对4种矿物蚀变信息提取模型进行参数寻优计算，以准确率为优化目标函数的适应度值，经多次迭代计算，分别得到各矿物识别模型在准确率最高时对应的最佳参数组合（表1）。

表1　极端随机树算法关键参数寻优结果

矿物种类	准确率	树深	决策树数量
褐铁矿	0.970	16	21
绿泥石	0.944	26	91
方解石	0.956	7	81
高岭土	0.782	5	81

利用主成分分析得到的特征波段数据集，基于寻优计算得到的最佳参数组合，分别构建4种矿物的提取模型，提取研究区矿化蚀变信息，结果见图2。

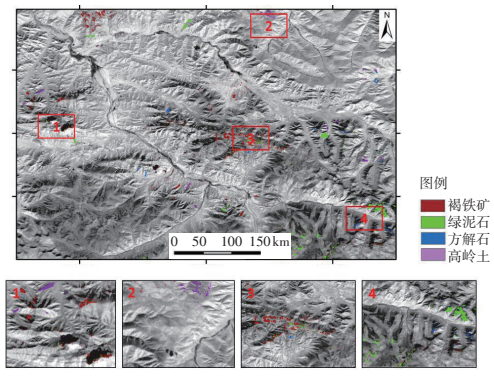


图2　基于极端随机树算法的蚀变信息提取结果

3　蚀变矿物提取效果评价

3.1　区域地质特征定性分析

为验证极端随机树算法在研究区内矿化蚀变信息的提取效果，将提取结果分别与区内断裂构造和已知矿床（点）分布进行空间叠加分析（图3），可以看出，蚀变矿物主要分布在构造行迹旁侧或其与次级构造交汇处，少量散布于主要构造行迹与其他构造的交切地段，与断裂构造的空间展布趋势具有很高的相似性；模型提取的蚀变信息主要以西北—东南向分布于地质构造带两侧的岩体中，且在金属矿床（点）周边地区也存在大量的蚀变信息；在研究区东南部模型同样提取出大量不同类型的蚀变信息，此处虽无已知矿（床）点分布，但成矿地质环境较优越，受断裂控制明显，三叠世洪水川组和二叠–石炭世浩特洛哇组的碳酸盐岩地层为该地段的主要地层，中志留世和中奥陶世的花岗闪长岩为其主要侵入岩体，因此该地段形成热液型金矿床的概率较大，可作为以后矿产勘查的重点地区。

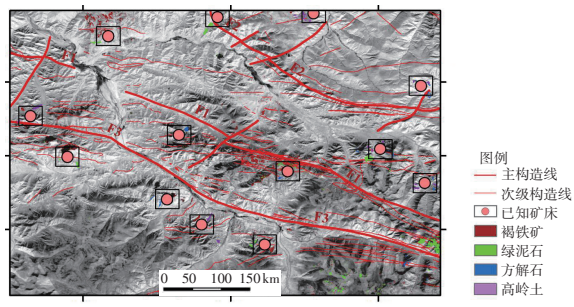


图3　基于地质特征的空间定性分析结果

注：F1为北界断裂；F2为狼日扎岗—龙里—尕之麻断裂；F3为东昆中断裂。

3.2　蚀变信息提取精度定量评价

为对比分析极端随机树算法的蚀变信息提取效果，本文分别利用基于光谱特征的混合调谐匹配滤波模型（MTMF）和支持向量机模型进行蚀变信息提取的对比实验[21-22]。支持向量机模型与极端随机树模型的构建过程基本一致，同样采用光谱角匹配技术获取样本数据集来完成模型的训练。MTMF模型首先基于MNF变换方法消除噪声影响，获取信息丰富波段；再采用MNF逆变换方法进行光谱重建，得到富含主要信息的地物真实反射率数值；然后利用已有样本数据，经均值计算得到端元平均光谱，并根据波谱MNF变换方法获取参考光谱反射率曲线，以此为基准与影像MNF变换结果进行滤波匹配；最后得到矿化蚀变信息提取结果。将3种模型提取的矿化蚀变信息分布与区内已知的13个矿床（点）进行空间叠加分析，采用吻合度指标来评价模型提取效果，数值越高，代表蚀变信息识别精度越好；反之，则越差，数学表达式为：

$$WD=\frac{L_{\text{in}}}{L_{\text{total}}}\times 100\% \tag{2}$$

式中，L_{in}为已知矿床（点）落在提取的矿化蚀变区域的个数；L_{total}为研究区已知矿床（点）总数。

计算得到的吻合度情况见表2，可以看出，极端

随机树模型的识别效果最好，各类矿物信息的吻合度均在0.46以上，其中褐铁矿识别效果最佳，其分布范围与8个已知矿点空间位置吻合，体现了极端随机树模型在稳定性和识别精度方面的优越性；MTMF模型的识别效果最差，方解石、高岭土的吻合度均低于0.4；支持向量机模型提取的各类矿物吻合度差异较大，除绿泥石外，褐铁矿、方解石、高岭土的吻合度均高于MTMF模型，说明基于机器学习算法构建的信息识别模型提取效果优于传统的光谱匹配算法。

表2 矿床吻合度计算结果

矿物种类	MTMF	支持向量机	极端随机树
褐铁矿	0.538	0.609	0.615
绿泥石	0.538	0.514	0.538
方解石	0.231	0.339	0.462
高岭土	0.308	0.455	0.462

4 结 语

本文利用ZY1–02D高光谱遥感数据，基于极端随机树算法构建矿化蚀变信息提取模型，对青海沟里地区典型矿物的提取进行了研究，并结合区内已知地质矿产资料对提取效果进行了评价。结果表明，结合光谱角和极端随机树算法的蚀变矿物提取模型对于各类矿物的提取效果明显优于MTMF模型和支持向量机模型，且提取的矿化蚀变信息分布范围与区内断裂构造展布特征以及已知矿床（点）的分布规律基本吻合，蚀变矿物的总体识别效果较好，可靠性较高，证明了机器学习算法在高光谱蚀变信息提取工作中的优越性，基于集成学习思想的极端随机树模型可作为矿产资源开发的有效技术手段加以利用。此外，极端随机树算法对于关键建模参数响应敏感，本文利用网格搜索—交叉验证方法，以准确率为适应度函数进行简单参数寻优，但该方法运算效率低、收敛速度过慢，在后续的研究工作中可引入群智能优化算法，用以探求不同参数寻优方法对蚀变信息提取结果的影响程度，提升矿化蚀变信息的提取精度。

参考文献

[1] 罗勇军，罗麟. 国产铁矿和进口铁矿价差扩大的原因探讨[J]. 现代矿业，2019，35(5):5-8

[2] 张燕斌，杜健民，王圆，等. 基于无人机高光谱遥感和3D-ResNet的荒漠草原地物分类[J]. 中国农机化学报，2022，43(4):66-73

[3] 成永生，周瑶. 土壤重金属高光谱遥感定量监测研究进展与趋势[J]. 中国有色金属学报，2021，31(11):3 450-3 467

[4] 张楠楠，周可法，孙莉，等. 遥感蚀变地质特征提取研究:以新疆康古尔地区为例[J]. 新疆地质，2008，26(1):39-42

[5] 薛云，戴塔根，邹林，等. 基于支持向量机的矿化蚀变信息提取研究:以青海黄南州吉地地区为例[J]. 地质调查与研究，2007，30(4):315-320

[6] 赵杰，周可法，崔世超，等. 联合光谱角匹配与随机森林的蚀变信息提取[J]. 遥感信息，2017，32(6):51-55

[7] Halotel J, Demyanov V, Gardiner A. Value of Geologically Derived Features in Machine Learning Facies Classification[J]. Mathematical Geosciences，2019，52(1):5-29

[8] 李廷伟，张莽，赵文文，等. 面向稀薄流非线性本构预测的机器学习方法[J]. 航空学报，2021，42(4):325-337

[9] 韦良芳. 基于极端随机树与Logistic回归算法的网贷平台个人信用评估模型的比较研究[D]. 济南:山东大学，2020

[10] 金康荣，於东军. 基于加权朴素贝叶斯分类器和极端随机树的蛋白质接触图预测[J]. 南京航空航天大学学报，2018，50(5):619-628

[11] 王彩艳. 基于ZY-3影像的北京市六环以内城市用地信息提取及特征分析[D]. 泰安:山东农业大学，2015

[12] 龚建周，陈健飞，刘彦随. E0-1 Hyperion高光谱影像的FLAASH大气校正与评价[J]. 广州大学学报(自然科学版)，2011，10(5):69-75

[13] 悦永峥，张廷斌，易桂花，等. 色那地区高光谱数据蚀变矿物提取及特征分析[J]. 地理空间信息，2016，14(3):18-21

[14] 吴艳，何政伟，赵银兵，等. 中高植被区斑岩型铜矿床矿化蚀变提取[J]. 地理空间信息，2015，13(3):93-95

[15] 占燕婷，吴柯，徐宏根，等. 联合光谱角与组合特征参数的高光谱影像分类[J]. 遥感信息，2021，36(6):140-146

[16] 高冬阳，龙华保，吴双卿，等. 基于光谱角与光谱距离的目标识别方法[J]. 大气与环境光学学报，2020，15(5):393-400

[17] Geurts P, Ernst D, Wehenkel L. Extremely Randomized Trees[J]. Machine Learning，2006，63(1):3-42

[18] 袁帅，余伟，余放，等. 面向不平衡数据集分类的离散高维空间距离采样和极端随机树算法[J]. 计算机应用与软件，2020，37(7):194-199

[19] Barrett B, Nitze I, Green S, et al. Assessment of Multi-temporal, Multi-sensor Radar and Ancillary Spatial Data for Grasslands Monitoring in Ireland Using Machine Learning Approaches[J]. Remote Sensing of Environment, 2014, 152:109-124

[20] Govindaraj R G, Subramaniyam S, Manavalan B. Extremely-randomized-tree-based Prediction of N6-Methyl adenosine Sites in Saccharomyces Cerevisiae[J]. Current Genomics，2020，21(1):26-33

[21] 田丰，董丽娜，杨苏明，等. 混合矿物组合光谱在蚀变矿物填图中的应用:以云南香格里拉地区Hyperion数据蚀变矿物填图为例[J]. 地质与勘探，2010，46(2):331-337

[22] 邹琪，何月顺，杨希，等. 基于集成学习的测井岩性识别模型的构建[J]. 智能计算机与应用，2020，10(3):91-94