

铁矿粉中全铁含量的 SFIM-RFR 高光谱预测模型

高伟¹, 杨可明^{1*}, 李孟倩², 李艳茹¹, 韩倩倩¹

1. 中国矿业大学(北京)地球科学与测绘工程学院, 北京 100083

2. 华北理工大学, 河北唐山 063210

摘要 铁矿是全球储量最高的金属矿产之一。全铁含量是评价铁矿石、铁矿粉品质的重要指标,在铁矿开采、矿石精选、矿粉冶炼等生产环节中有特殊意义。传统的铁矿粉全铁含量化学分析方法存在耗时久、操作复杂、污染严重等优点,因此,探寻一种快速、有效、无污染的检测方法越来越成为矿山环境的研究热点。高光谱技术具有光谱分辨率高、曲线连续、无损伤、无污染、可对物质特征或成分进行精确探测等特点。使用铁矿粉高光谱数据,通过建立用于光谱特征筛选的光谱特征重要性评分(SFIM)指标,并结合随机森林回归(RFR)方法构建铁矿粉全铁含量预测的 SFIM-RFR 模型。以河北省阳原县三义庄铁矿为研究区,于2018年11月与2019年3月在研究区收集铁精粉、铁尾砂原料,分别制作第一批次的训练组和验证组铁矿粉试样以及第二批次的二次验证组铁矿粉试样,并使用 ASD Field Spec4 型光谱仪测量试样的光谱反射率;然后使用第一批次的训练组光谱数据训练 SFIM-RFR 模型,对第一批次的验证组样本的全铁含量进行预测,同时采用常规 RFR、线性回归(LR)预测模型来对比分析铁矿粉样本全铁含量预测结果;最后使用二次验证组光谱数据检验多模型鲁棒性。结果表明:SFIM-RFR, RFR 和 LR 模型全铁含量预测结果与2018年11月采集的验证组样本全铁含量真实值的确定系数(R-Square)分别为0.991 8, 0.988 4 和 0.898 7, 均方根误差(RMSE)分别为0.016 9, 0.020 1 和 0.059 6, 多模型预测效果总体较好, SFIM-RFR 模型预测结果误差最小,说明了 SFIM-RFR 模型用于预测铁矿粉中全铁含量的可行性和有效性,且 SFIM-RFR 模型预测效果优于常规的预测模型;SFIM-RFR, RFR 和 LR 模型全铁含量预测结果与2019年3月采集的二次验证组样本全铁含量真实值的 R-square 分别为0.976 8, 0.974 5 和 0.914 0, RMSE 分别为0.034 6, 0.036 2 和 0.071 9, 证明了 SFIM-RFR 模型的预测效果较为理想且鲁棒性很强。

关键词 高光谱; 铁矿粉全铁含量; 预测模型; 光谱特征重要性评分; 随机森林回归

中图分类号: TP75 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2020)08-2546-06

引言

铁是人类最早发现、用途最广泛、用量最大的一种具有战略地位的金属。铁矿粉中全铁含量的测定在铁矿开采、铁矿石精选、铁矿粉冶炼等环节及铁矿粉品质评价等方面有重要意义,目前国内外使用最广泛的铁矿粉全铁含量测定方法为化学分析法^[1-2],但此类方法存在时间长、药品昂贵、化验废液污染环境等诸多缺点。高光谱遥感数据具有光谱分辨率高、谱线连续、隐含信息丰富等特点^[3],已广泛应用于物质含量预测与丰度反演,如基于高光谱数据的变换等处理,可采用线性回归(linear regression, LR)、多元逐步回归

(stepwise multiple linear regression, SMLR)、偏最小二乘回归(partial least-squares regression, PLSR)、支持向量机(support vector machine, SVM)等方法预测植物叶绿素含量^[4-5];采用模糊识别(fuzzy recognition, FR)、SMLR、PLSR等方法预测土壤中有机碳含量、有机质含量和重金属元素含量^[6-8];也有学者使用高光谱数据结合 PLSR 和 LR 等回归分析方法对土壤、植物叶片中全铁及其他形态铁的含量进行预测研究^[9-10]。

国内外学者在使用光谱数据进行铁矿物质全铁含量预测方面取得了一定成果。Yangmin G^[11]等将稀疏偏最小二乘(hybrid sparse partial least-squares, SPLS)与最小二乘支持向量机(least-squares support vector machine, LS-SVM)结

收稿日期: 2019-12-09, 修订日期: 2020-03-19

基金项目: 国家自然科学基金项目(41971401)资助

作者简介: 高伟, 1997年生, 中国矿业大学(北京)硕士研究生

e-mail: gw970606@163.com

* 通讯联系人 e-mail: ykm69@163.com

合,建立了分析铁矿石纯度的 SPLS-LS-SVM 模型,模型性能较传统的 SPLS 模型和 LS-SVM 模型更好;李颖娜^[12]等建立了基于反向传播(back propagation, BP)和径向基函数(radial basis function, RBF)的集成神经网络多物相铁矿石全铁含量预测模型,实现了铁矿石相分类与全铁含量预测;何群^[13]等建立了条带状铁建造铁矿石全铁含量的 PLSR 预测模型,全铁含量的预测误差为 3.43%。现有的全铁含量预测模型存在一定的局限性,如预测范围有限,偏重对铁矿石全铁含量的预测等。因此,本文以河北省三义庄铁矿为研究区,基于在研究区提取的铁精粉和铁尾砂原料,制备铁矿粉样本,对其进行全铁含量测定和光谱数据采集,将光谱特征重要性评分(spectral feature importance measure, SFIM)与随机森林回归(random forest regression, RFR)相结合,构建高光谱遥感的铁矿粉全铁含量 SFIM-RFR 预测模型;同时,通过比较分析 SFIM-RFR 模型与常规 RFR、LR 模型的预测结果以及样品化学分析的测定结果,验证 SFIM-RFR 模型的预测精度与鲁棒性,探索快速、有效、无污染检测铁矿粉全铁含量的新方法。

1 实验部分

1.1 随机森林回归(RFR)

随机森林是聚合多棵分类与回归树(classification and regression tree, CART)的随机抽样判别模型,随机森林进行回归预测时,CART 树的返回结果为离散值,取所有 CART 树返回结果的平均值为最终输出结果^[14]。

1.2 光谱特征重要性评分(SFIM)

变量重要性评分(variable importance measure, VIM)是随机森林中的指标参数,在其原理基础上构建光谱特征重要性评分(SFIM)来表征各特征波段光谱数据对随机森林预测结果的影响程度。每个特征波段(x_j)的 SFIM_{*j*} 计算依据是 x_j 在随机森林中每棵 CART 树(f_k)处未参与抽样数据预测误差的均方差, x_j 在 f_k 处的 SFIM_{*jk*} 为

$$\text{SFIM}_{jk} = \left| \frac{\sum_{p=1}^{m_k} (y_p - y_{p1})^2}{m_k} - \frac{\sum_{p=1}^{m_k} (y_p - y_{p2})^2}{m_k} \right| \quad (1)$$

式(1)中, m_k 是第 k 棵 CART 树处的预测例数, y_p 是参与抽样数据产生的第 p 个预测结果, y_{p1} 是未参与抽样数据产生的第 p 个预测结果, y_{p2} 是将 x_j 随机置换后未参与抽样数据产生的第 p 个预测结果。所以, x_j 的 SFIM_{*j*} 为

$$\text{SFIM}_j = \frac{\sum_{k=1}^n \text{SFIM}_{jk}}{n} \quad (2)$$

式(2)中,SFIM_{*jk*} 为 x_j 在 f_k 处的光谱特征重要性评分, n 为随机森林中 CART 树的数量。

1.3 SFIM-RFR 预测模型

将 SFIM 与 RFR 相结合,建立 SFIM-RFR 预测模型。依据光谱数据各特征波段的 SFIM 对光谱数据进行遴选,由 SFIM 高的特征波段构成优势光谱数据集,使用优势光谱数据集进行训练、预测。SFIM-RFR 模型适用于多特征目标的

预测,同时减少了无关光谱波段信息的影响,较常规模型具有优势。

1.4 研究区概况

实验研究区为三义庄铁矿,三义庄铁矿位于河北省阳原县化稍营镇,处于辽西—冀北地区华北地台北缘大型多金属成矿带的西部,地理位置如图 1 所示,该矿区的矿体形态、矿化组合特点等方面较为复杂,矿石组分各有不同。

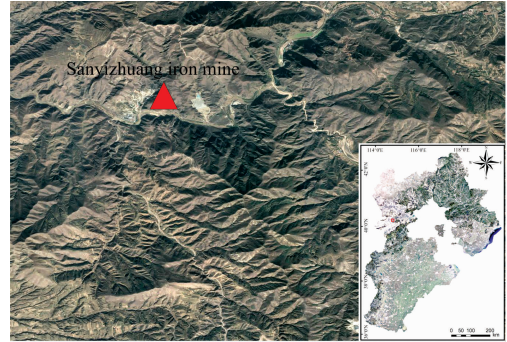


图 1 三义庄铁矿地理位置示意图

Fig. 1 Map of geographical location of Sanyizhuang iron mine

1.5 样本采集与数据处理

2018 年 11 月、2019 年 3 月从河北省三义庄铁矿分别采集了两批次的铁精粉和铁尾砂原料,在实验室对铁精粉、铁尾砂原料做干燥、研磨、过筛等处理,化验测定原料的全铁含量真实值。原料的全铁含量分别为:0.7%(2018 年 11 月铁尾砂)、67.34%(2018 年 11 月铁精粉)、2.67%(2019 年 3 月铁尾砂)、64.34%(2019 年 3 月铁精粉)。使用 2018 年 11 月批次的铁精粉、铁尾砂原料按不同质比配制 28 组铁矿粉样本,使用 2019 年 3 月批次的铁精粉、铁尾砂原料按不同质比配制 7 组铁矿粉样本,计算得到每组样本的全铁含量,再将每组样本搅拌均匀后分成 5 份平行试样(A, B, C, D 和 E),共计 35 组 175 个试样。

依据验证组占比 25% 的原则,将 2018 年 11 月批次原料制备的 28 组样本划分为 21 组训练组样本和 7 组验证组样本,将 2019 年 3 月原料制备的 7 组样本划分为二次验证组。依据样本全铁含量对试样进行标记,训练组 21 组样本的标记情况如表 1 所示,验证组 7 组样本的标记情况如表 2 所示,二次验证 7 组样本的标记情况如表 3 所示。

光谱采集仪器为 ASD Field Spec4 型光谱仪,该仪器的探测器由检测波长范围为 350~1 000, 1 001~1 800 和 1 801~2 500 nm 的三个传感器拼接而成。测量光谱时,将样本平铺在黑色不反光纸上,保持样本表面平整,光源为光谱仪探头内置光源,将探头底部垂直贴近样本,确保不漏光。对每种样本的 A, B, C, D 和 E 平行试样分别进行 5 次光谱数据采集,每组样本数据采集后重新进行标准化白板校正,共获得 875 次光谱测量结果。

为了消除 ASD Field Spec4 型光谱仪三个传感器拼接造成的误差影响,对采集的光谱进行拼接校正(splice correction, SC)。使用 Savitzky-Golay(SG)卷积平滑算法对光谱曲线进行处理;由于仪器自身结构的影响,所获取的光谱数据

表 1 训练组样本全铁含量及标记

Table 1 Total iron contents and labeling of training group samples

全铁含量/%	标记	全铁含量/%	标记
0.70	TFe(0.70)	39.99	TFe(39.99)
5.02	TFe(5.02)	42.63	TFe(42.63)
7.43	TFe(7.43)	45.00	TFe(45.00)
10.10	TFe(10.10)	47.50	TFe(47.50)
12.66	TFe(12.66)	54.97	TFe(54.97)
17.54	TFe(17.54)	57.50	TFe(57.50)
20.16	TFe(20.16)	62.50	TFe(62.50)
24.91	TFe(24.91)	63.68	TFe(63.68)
27.57	TFe(27.57)	64.25	TFe(64.25)
30.05	TFe(30.05)	67.34	TFe(67.34)
37.72	TFe(37.72)		

在“首”、“尾”存在较大的噪声,可对边缘光谱以 50 nm 为限值进行剔除,保留 400~2 450 nm 的光谱信息。剔除各组样本的 A, B, C, D 和 E 平行试样内光谱数据的异常值后,求取各组内剩余试样的均值光谱,即获得 35 条不同全铁含量

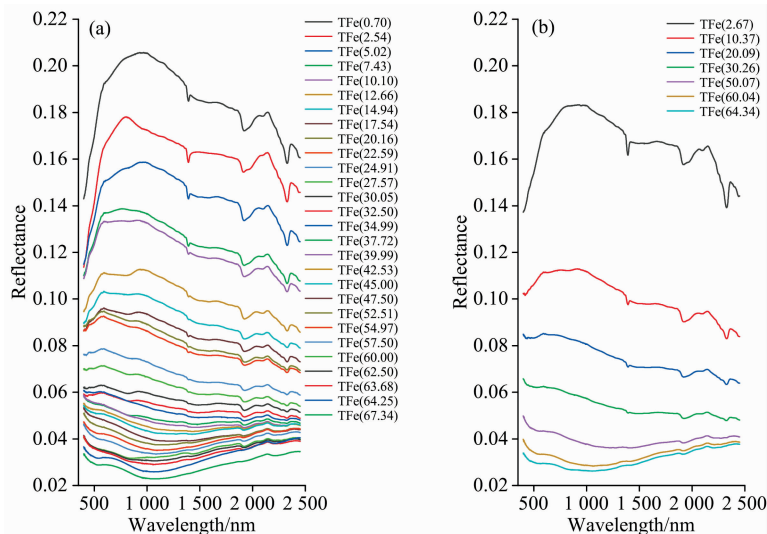


图 2 铁矿粉样本光谱曲线预处理结果图

(a): 2018 年 11 月批次铁矿粉样本光谱曲线预处理结果; (b): 2019 年 3 月批次铁矿粉样本光谱曲线预处理结果

Fig. 2 Spectral curve processing results of iron ore powder samples

(a): Samples collected in November 2018; (b): Samples collected in March 2019

1.6 模型构建与预测

1.6.1 SFIM-RFR 模型构建与预测

为了检验 SFIM-RFR 模型的预测效果,将训练组数据和验证组数据输入 SFIM-RFR 模型,训练模型并对验证组铁矿粉样本全铁含量进行预测,得到多次预测的结果平均值,以及平均确定系数 (R-Square) 和平均方根误差 (RMSE)。SFIM-RFR 预测模型的 α 次预测过程为:

- (1) 输入训练组数据和验证组数据;
- (2) 设置 SFIM-RFR 模型中 CART 树的数量;
- (3) 使用训练组数据训练 SFIM-RFR 模型,并对验证组数据进行预测;

的铁矿粉样本光谱曲线,如图 2 所示。

表 2 验证组样本全铁含量及标记

Table 2 Total iron contents and labeling of testing group samples

全铁含量/%	标记	全铁含量/%	标记
2.54	TFe(2.54)	34.99	TFe(34.99)
14.94	TFe(14.94)	52.51	TFe(52.51)
22.59	TFe(22.59)	60.00	TFe(60.00)
32.50	TFe(32.50)		

表 3 二次验证组样本全铁含量及标记

Table 3 Total iron contents and labeling of second testing group samples

全铁含量/%	标记	全铁含量/%	标记
2.67	TFe(2.67)	50.07	TFe(50.07)
10.37	TFe(10.37)	60.04	TFe(60.04)
20.09	TFe(20.09)	64.34	TFe(64.34)
30.26	TFe(30.26)		

- (4) 计算各特征波段的 SFIM;
- (5) 将 SFIM 最大的波段选入新特征波段集;
- (6) 若新特征波段集中的波段数不等于阈值 γ , 转到(3); 若等于, 转到(7);
- (7) 依据新特征波段集组建新训练组数据、新验证组数据;
- (8) 使用新训练组数据训练 SFIM-RFR 模型, 并对新验证组数据进行预测;
- (9) 依据预测结果, 计算 R-Square 和 RMSE;
- (10) 设 $i=i+1$ (初始 i 为 0);
- (11) 若 i 不等于阈值 α , 转到(8); 若 i 等于阈值 α , 转到

(12);

(12)输出 α 次预测结果、R-Square 和 RMSE 的平均值。

1.6.2 RFR 模型构建与预测

为与 SFIM-RFR 模型进行对比, 结合 RFR 原理, 建立 RFR 全铁含量预测模型, 将训练组数据和验证组数据输入 RFR 模型, 训练模型并对验证组铁矿粉样本全铁含量进行预测, 得到多次预测的结果平均值, 以及平均 R-Square 和平均 RMSE。RFR 模型的 β 次预测过程为:

- (1)输入训练组数据和验证组数据;
- (2)设置 RFR 模型中 CART 树的数量;
- (3)使用训练组数据训练 RFR 模型, 并对验证组数据进行预测;
- (4)依据预测结果, 计算 R-Square 和 RMSE;
- (5)设 $i=i+1$ (初始 i 为 0);
- (6)若 i 不等于阈值 β , 转到(3); 若 i 等于阈值 β , 转到(7);
- (7)输出 β 次预测结果、R-Square 和 RMSE 的平均值。

1.6.3 LR 模型构建与预测

LR 是最常用的回归分析方法之一, 故基于 LR 建立全铁含量预测的 LR 模型, 将光谱反射值与样本全铁含量相关系数最大的波段作为模型的输入值。计算铁矿粉样本各波段光谱反射值与样本全铁含量之间的相关系数, 结果显示, 铁矿粉样本的光谱反射值与样本全铁含量之间的相关性整体较

强, 相关系数绝对值最大的波段为 400 nm 处, 相关系数达到 -0.9535 , 呈显著负相关。使用与样本全铁含量相关系数绝对值最大的 400 nm 处的光谱数据训练 LR 模型, 表达式为 $y = -689.9x_{400} + 83.87$, 对验证组铁矿粉样本全铁含量进行预测, 得到预测结果、R-Square 和 RMSE。

2 结果与讨论

2.1 模型预测结果与精度分析

SFIM-RFR 模型、RFR 模型、LR 模型对验证组铁矿粉样本全铁含量的预测结果与真实值的差异如图 3 所示, 由图 3(a)与图 3(b)所示 SFIM-RFR 模型和 RFR 模型得出的预测值与真实值的差异较小, 而图 3(c)所示 LR 模型得出的预测值与真实值的差异略大。3 个模型得到的样本全铁含量预测值与真实值的 R-Square 与 RMSE 见表 4, SFIM-RFR 模型的 R-Square 为 0.9918, RMSE 为 0.0169; RFR 模型的 R-Square 为 0.9884, RMSE 为 0.0201; LR 模型的 R-Square 为 0.8987, RMSE 为 0.0596; 可见, SFIM-RFR 模型的预测结果最好, RFR 模型的预测结果较好, LR 模型的预测结果较差。总体来说, 利用铁矿粉高光谱数据建立的 LR 模型、RFR 模型、SFIM-RFR 模型对铁矿粉的全铁含量都有一定的预测能力, 其中 SFIM-RFR 模型的预测精度最好。

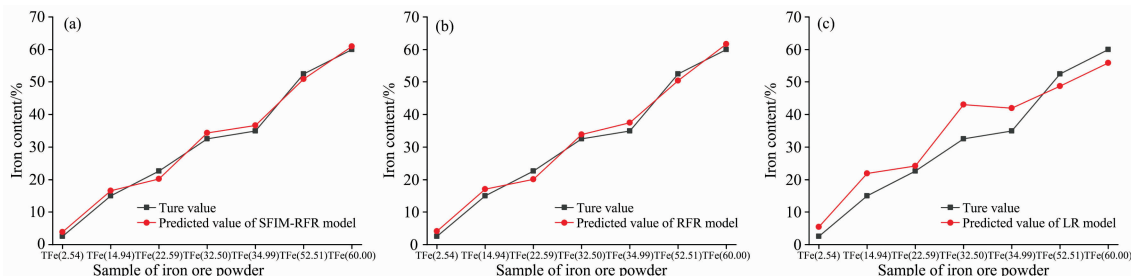


图 3 全铁含量的 SFIM-RFR、RFR 与 LR 模型预测值与真实值差异图

(a): SFIM-RFR 模型; (b): RFR 模型; (c): LR 模型

Fig. 3 The difference between the predicted value and the true value of total iron contents

(a): SFIM-RFR model; (b): RFR model; (c): LR model

表 4 基于训练组和验证组数据的模型预测精度表

Table 4 Prediction accuracy of the model based on the training and testing data

Model	R-Square	RMSE
SFIM-RFR	0.9918	0.0169
RFR	0.9884	0.0201
LR	0.8987	0.0596

2.2 模型鲁棒性检验

为了检验模型的鲁棒性, 使用二次验证组数据对 SFIM-RFR 模型、RFR 模型、LR 模型进行检验。3 个模型得出的样本全铁含量预测值与真实值的差异如图 4 所示, 由图 4(a)与图 4(b)所示 SFIM-RFR 模型和 RFR 模型得出的预测值与

真实值的差异较小, 而图 4(c)所示 LR 模型得出的预测值与真实值的差异较大。3 个模型得到的样本全铁含量预测值与真实值的 R-Square 与 RMSE 见表 5, SFIM-RFR 模型的 R-Square 为 0.9768, RMSE 为 0.0346, 预测值与真实值较为接近, 预测效果较好, 模型鲁棒性较强; RFR 模型的 R-Square 为 0.9745, RMSE 为 0.0362, 预测值与真实值差异略大, 预测效果不甚理想; LR 模型的 R-Square 为 0.9140, RMSE 为 0.0719, 预测值与真实值差异较大, 预测效果较差。总的来说, SFIM-RFR 模型的预测结果较为理想, 与样本全铁含量真实值较为接近, 模型预测能力较为稳定, 鲁棒性较强, 可以在一定精度范围内实现通过铁矿粉样本高光谱数据对样本全铁含量进行预测。

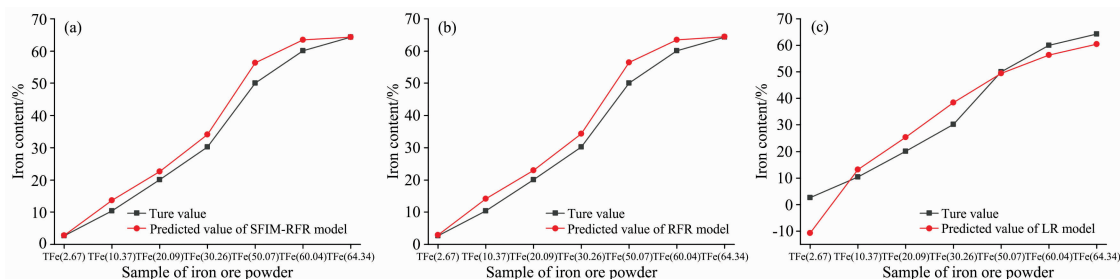


图 4 全铁含量的 SFIM-RFR、RFR 与 LR 模型预测值与二次验证组数据差异图

(a): SFIM-RFR 模型; (b): RFR 模型; (c): LR 模型

Fig. 4 The difference between the predicted value and the second testing data of total iron contents

(a): SFIM-RFR model; (b): RFR model; (c): LR model

表 5 基于二次验证组数据的模型检验精度表

Table 5 Test accuracy of the model based on the second testing data

Model	R-Square	RMSE
SFIM-RFR	0.976 8	0.034 6
RFR	0.974 5	0.036 2
LR	0.914 0	0.071 9

3 结 论

SFIM-RFR 预测模型可利用铁矿粉高光谱数据对铁矿粉的全铁含量进行有效预测, 通过与常规 RFR、LR 模型预测结果进行比较, 可以发现:

(1) 使用 2018 年 11 月采集的训练组数据与验证组数据进行预测, 结果说明 SFIM-RFR 模型对全铁含量的预测能力最强, R-Square 为 0.991 8, RMSE 为 0.016 9; RFR 模型的

预测能力次之, R-Square 为 0.988 4, RMSE 为 0.020 1; 而 LR 模型的预测能力最差, R-Square 为 0.898 7, RMSE 为 0.059 6。3 个模型都能基于铁矿粉高光谱数据对全铁含量进行一定程度的预测, 证明了使用高光谱数据进行铁矿粉全铁含量预测的可行性, 其中 SFIM-RFR 模型的预测效果最好。

(2) 使用 2019 年 3 月采集的二次验证组数据对 SFIM-RFR 模型、RFR 模型、LR 模型的鲁棒性与普适性进行检验, 结果显示 LR 模型的 R-Square 为 0.914 0, RMSE 为 0.071 9, 预测值与真实值差异较大, 模型鲁棒性较差; RFR 模型的 R-Square 为 0.974 5, RMSE 为 0.036 2, 预测值与真实值差异略大, 模型鲁棒性一般; SFIM-RFR 模型的 R-Square 为 0.976 8, RMSE 为 0.034 6, 预测值与真实值最为接近, 预测效果最好, 模型预测能力最为稳定, 鲁棒性最强, 可以在一定精度范围内实现依据铁矿粉高光谱数据的全铁含量预测。

References

- [1] Hu H, Tang Y, Ying H, et al. *Talanta*, 2014, 125: 425.
- [2] MING Hong, REN Li-ping, QIN Ye-qiong, et al(闵红, 任丽萍, 秦晔琼, 等). *Metallurgical Analysis(冶金分析)*, 2014, 34(4): 21.
- [3] TONG Qing-xi, ZHANG Bing, ZHANG Li-fu(童庆禧, 张兵, 张立福). *Journal of Remote Sensing(遥感学报)*, 2016, 20(5): 689.
- [4] DOU Zhi-guo, CUI Li-juan, WU Gao-jie, et al(窦志国, 崔丽娟, 武高洁, 等). *Journal of Ecology(生态学杂志)*, 2018, 37(10): 3163.
- [5] ZHANG Zhuo-ran, CHANG Qing-rui, ZHANG Ting-long, et al(张卓然, 常庆瑞, 张廷龙, 等). *Journal of Northwest A&F University · Natural Science Edition(西北农林科技大学学报·自然科学版)*, 2018, 46(11): 1.
- [6] SUN Wen-juan, LI Xin-ju(孙问娟, 李新举). *Journal of Soil and Water Conservation(水土保持学报)*, 2018, 32(5): 346.
- [7] LI Xi-can, ZHAO Geng-xing, CHEN Hang-yan, et al(李西灿, 赵庚星, 陈红艳, 等). *Journal of Geomatics Science and Technology(测绘科学技术学报)*, 2014, 31(6): 593.
- [8] SONG Ting-ting, FU Xiu-li, CHEN Yu, et al(宋婷婷, 付秀丽, 陈玉, 等). *Remote Sensing Technology and Application(遥感技术与应用)*, 2018, 33(1): 88.
- [9] Liu Y, Chen Y. *International Journal of Remote Sensing*, 2012, 33(18): 5954.
- [10] Basayigit L, Dedeoglu M, Akgul H. *Turkish Journal of Agriculture and Forestry*, 2015, 39(1): 123.
- [11] Guo Y M, Guo L B, Hao Z Q, et al. *Journal of Analytical Atomic Spectrometry*, 2018, 33(8): 1330.
- [12] LI Ying-na, XU Zhi-bin(李颖娜, 徐志彬). *Metallurgical Analysis(冶金分析)*, 2019, 39(1): 35.
- [13] HE Qun, WANG Dong, LIU Shan-jun, et al(何群, 王东, 刘善军, 等). *Metal Mine(金属矿山)*, 2018, 12: 140.
- [14] Breiman L. *Machine Learning*, 2001, 45(1): 5.

Hyperspectral SFIM-RFR Model on Predicting the Total Iron Contents of Iron Ore Powders

GAO Wei¹, YANG Ke-ming^{1*}, LI Meng-qian², LI Yan-ru¹, HAN Qian-qian¹

1. College of Geoscience and Surveying Engineering, China University of Mining & Technology (Beijing), Beijing 100083, China

2. North China University of Science and Technology, Tangshan 063210, China

Abstract Iron ore is one of the most abundant metallic minerals in the world. Total iron contents is an important index to evaluate the quality of iron ore and iron ore powder, and it has a special significance in iron ore mining, ore dressing, ore smelting and other production links. The traditional chemical methods have the disadvantages of a time-consuming, complex operation, seriously pollution. Therefore, exploring a new method of rapid, effective and pollution-free detection has become a hot spot in mine environment research. Hyperspectral technology has the characteristics of high spectral resolution, continuous curve, no damage, no pollution and accurate detection of characteristics or components of materials. The purpose of this paper is to establish a data evaluation index of spectral feature importance measures (SFIM) and to screen spectral features based on the hyperspectral data of iron ore powder, and then combined with random forest regression (RFR) to establish the SFIM-RFR prediction model and predict the total iron contents of iron ore powder. First, taking Sanyizhuang iron mine in Yangyuan county, Hebei province as a research object, based on the iron concentrate and iron powder tail collected in the research area in November 2018 and March 2019, the first batch of iron ore powder samples in the training group and the testing group and the second batch of iron ore powder samples in the second testing group were made respectively. Spectral data of samples were measured by the ASD Field Spec4 spectrometer. Then, spectral data of the first batch of training group were used in the SFIM-RFR model training, and the total iron contents in the samples of the first batch of the testing group were predicted. Meanwhile, conventional methods, including RFR and linear regression (LR) prediction model, were used to compare and analyze the predicted results of total iron contents in iron ore powder samples. Finally, the spectral data of the second testing group were used to validate the robustness of the multi-model. The results show that the R-Square values of prediction results of total iron contents by the SFIM-RFR, RFR and LR models were 0.991 8, 0.988 4, 0.898 7, and RMSE values are 0.016 9, 0.020 1, 0.059 6. The results of multi-model prediction overall are good, and the SFIM-RFR model has the minimum error, which indicates the feasibility and effectiveness of this model in predicting the total iron contents of iron ore powder. Moreover, the prediction ability of SFIM-RFR model is better than that of conventional prediction models. The R-Square values of the prediction results of total iron contents by the SFIM-RFR, RFR and LR models are 0.976 8, 0.974 5 and 0.914 0. The RMSE values are 0.034 6, 0.036 2 and 0.071 9, which proves that the prediction ability of the SFIM-RFR model is the best and the robustness of the prediction model is the best.

Keywords Hyperspectral; Total iron contents of iron ore powder; Prediction model; SFIM; Random forest regression

(Received Dec. 9, 2019; accepted Mar. 19, 2020)

* Corresponding author