

可见光-近红外光谱的低品位斑岩型铜矿反演模型

毛亚纯^{1,2}, 丁瑞波^{1*}, 刘善军^{1,2}, 包妮沙^{1,2}

1. 东北大学资源与土木工程学院, 辽宁 沈阳 110819
2. 东北大学智慧矿山研究中心, 辽宁 沈阳 110819

摘要 目前, 国内外铜矿品位分析多以化学分析法为主, 但由于化学分析法存在成本高、时间长和污染物残留等缺点, 其相对配矿流程存在严重的滞后效应, 致使尾矿铜含量过高, 必然造成资源浪费。开展斑岩型铜矿可见光-近红外光谱特征与建模研究是解决这一问题的有效途径。以 121 个乌山斑岩型铜矿的化学分析与光谱测试数据为数据源, 分析了斑岩型铜矿可见光-近红外光谱特征, 以主成分分析法(PCA)、局部线性嵌入算法(LLE)两种降维算法对原始光谱数据进行了处理, 所降维数分别为 3 维和 5 维, 同时利用遗传算法(GA)对原始光谱数据进行了波段选择, 共选取了 467 个最佳波段。然后以 BP 神经网络为建模方法, 并分别以 92 个和 29 个斑岩型铜矿可见光-近红外光谱数据作为建模样本和测试样本, 建立了斑岩型铜矿可见光-近红外光谱的定量反演模型。利用原始数据所建模型的品位反演平均绝对误差为 0.104%, 利用主成分分析法、局部线性嵌入算法、遗传算法处理后的数据所建模型品位反演平均绝对误差分别为 0.110%, 0.093% 和 0.045%, 由此可见, 利用主成分分析法处理后的数据所建模型品位反演精度较差, 利用局部线性嵌入算法处理后的数据所建模型品位反演精度略有提高, 而利用遗传算法处理后的数据所建模型品位反演精度有明显提高。研究表明, 基于低品位斑岩型铜矿可见光-近红外光谱数据反演模型的品位分析具有一定的可行性, 为我国低品位斑岩型铜矿的品位快速检测提供了一种有效的手段。

关键词 斑岩型铜矿; 可见光-近红外光谱; 降维算法; 遗传算法; 反演模型

中图分类号: TP79 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2020)08-2474-05

引言

斑岩型铜矿是我国铜矿的主要矿种之一, 约占我国铜矿总储量的 41%^[1], 但品位普遍偏低, 平均品位仅为 0.55%^[2]。目前, 其品位分析的主要方法为化学分析法, 由于化学分析方法工艺复杂、化验周期较长, 相对配矿流程存在滞后效应, 因此难以适应精准配矿的要求。近年来, 随着机器学习及高光谱分析的快速发展, 基于高光谱定量反演建模方法被普遍应用于植被重金属胁迫、土壤污染和矿石品位分析等领域^[3-7]。

针对高光谱原始数据处理及定量反演建模等问题, 大量学者对高光谱数据进行了特征参数提取、敏感波段选取及数据降维处理等方法研究^[8-10], 这些方法有效的降低了高光谱数据的数据量及数据冗余, 为准确、快速建模奠定了基础。另外, 很多学者也对建模方法进行了深入研究, 传统的高光

谱数据建模方法多以偏最小二乘法为主, 随着机器学习的发展, 机器学习法被广泛应用于反演建模中。Liu 等利用 SMLR, PLSR 和 PCR 相结合的方法成功建立了不同土壤的近红外光谱与其有机碳之间的反演模型^[11]; Liang 等以小麦的高光谱数据为数据源, 并通过建立 FD-NDNI 及 FD-SRNI 两种新的高光谱指数, 利用随机森林方法对小麦的氮含量进行预测, 取得了较好的效果^[12]。虽然国内外学者对基于高光谱数据处理及建模方法进行了较为深入的研究, 但针对斑岩型铜矿的高光谱数据处理及反演建模方法研究相对较少。

以乌山斑岩型铜矿的化学分析与光谱测试数据为数据源, 并对其降维及波段选择处理, 同时以 BP 神经网络为建模方法建立了品位定量反演模型。研究表明, 可见光-近红外光谱在低品位斑岩型铜矿品位分析方面具有一定的可行性, 能够为我国斑岩型铜矿品位的快速分析提供一种有效的手段。

收稿日期: 2019-07-17, 修订日期: 2019-11-30

基金项目: 国家重点研发计划项目(2016YFC0801602), 国家自然科学基金项目(41371437)资助

作者简介: 毛亚纯, 1966 年生, 东北大学资源与土木工程学院副教授 e-mail: dbdxmyc@163.com

* 通讯联系人 e-mail: 1960446852@qq.com

1 实验部分

1.1 样品采集

样品采自中国黄金集团内蒙古乌山铜钼矿露天采场。在矿区不同地点采集大小适中的铜钼矿块状样本。将采集的样本经过筛选、研磨，共制成粉末状样本 121 件。

1.2 光谱测试及化验

使用美国 SVC HR-1024 便携式地物光谱仪对 121 件粉末状样品进行光谱测试。该仪器波段范围为 350~2 500 nm，通道数为 1 024，光谱精度优于±0.5 nm，光谱分辨率≤8.5 nm，最小积分时间为 1 s。为降低气溶胶及太阳辐射传播路径的影响，得到可靠的测试数据，实验在 10:00—14:00 进行。测量时要求天空晴朗无云，太阳高度角在 45°左右。为避免测量背景影响，将被测样品放置于边长为 5cm 的正方形黑色小盒中，测量时要求样品表面平整，光谱仪镜头垂直于样品观测面，采样积分时间设置为 2 s，视场角为 4°。为避免测试数据受方向性的影响，每个样品在测试过程中均水平旋转 3 次，每次旋转约 90°，每个角度测试 1 次，取 4 次测试的反射率均值绘制该样品的光谱曲线。

光谱测试结束后，将全部实验样品进行了化学成分测试，以此确定各个实验样品的铜含量。由化验结果得出，该次实验样本铜品位在 0.062%~0.782%之间，平均铜品位为 0.279%。低于我国的斑岩型铜矿平均品位。

2 结果与讨论

图 1 为全部实验样品的光谱曲线，光谱具有如下特征：

(1) 样品的光谱反射率大部分在 40%~70%之间。

(2) 350~600 nm 波段光谱曲线出现差异，呈现出两种变化趋势，一种为上升趋势，另一种变化平缓；600~2 500 nm 波段所有光谱曲线差别不大呈现较平缓的趋势。

(3) 实验样品中含有部分水(包括孔隙水和结晶水)，所有实验样品光谱于 1 400 与 1 900 nm 附近出现波谷，且该处光谱曲线毛刺较多，波动较大。

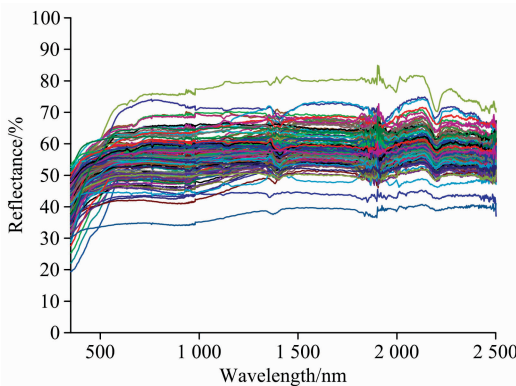


图 1 实验样品可见光-近红外光谱曲线
Fig. 1 Visible and near infrared spectra of experimental samples

(4) 所有光谱曲线于 760nm 处出现微弱波谷。

对上述现象进行分析发现，在 350~600 nm 间的光谱差异与样本铜含量有一定关系，上升快的样本铜含量均值远小于较为平缓的铜含量均值。

2.1 数据处理

2.1.1 基于降维算法的预处理

(1) 主成分分析法

主成分分析法^[13]是一种线性数据降维分析方法，采用主成分分析法对斑岩型铜矿原始光谱数据进行降维分析，设置累计贡献率为 95%，依据累计贡献率计算最终降维维度 d 。经处理后，原始数据被降为 3 维，各维主成分所占贡献率如图 2 所示。图中横坐标为各主成分，纵坐标为各主成分贡献率。

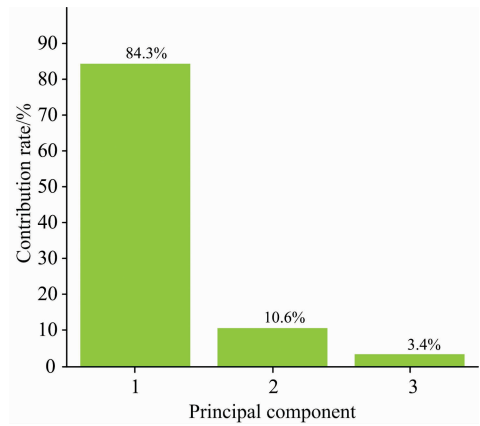


图 2 主成分分析结果

Fig. 2 Principal component analysis results

(2) 局部线性嵌入算法

高光谱数据在采集过程中受多种因素影响，往往包含大量的非线性因素。局部线性嵌入算法^[14]是一种非线性降维算法，用于对原始数据进行降维处理时，由于近邻点 k 与所降维度 d 的取值对最终降维结果有较大影响，因此设置该算法的近邻点 k 的取值为 1~100，并以欧氏距离寻找每个样本点的近邻点，同时参照主成分分析法的降维结果，设置局部线性嵌入算法降维维度 d 为 2~20 维，用最终的降维结果为数据源，采用 BP 神经网络算法反演铜含量，以反演结果的平均绝对误差为依据，确定了最邻近数 k 为 17，降维维度为 5 维。

2.1.2 基于遗传算法的最佳波段选取

遗传算法^[15]是一种以遗传机制和生物进化理论为基础的最优化并行随机搜索方法，它在建模自变量提取方面具有很大的优越性。遗传算法主要有三个基本操作：选择、交叉和变异。

使用遗传算法对原始数据进行处理时，首先将 121 个样本按照铜含量均匀选出训练集与测试集，其中训练集包含 92 个样品，测试集包含 29 个样品。然后利用遗传算法对 121 个样品的原始数据进行处理。根据原始数据维度，将遗传算法中染色体长度设置为 973，进化代数设置为 40。首先产生初始种群，随机生成相应的阈值和权值，对各个个体进行编码

处理并计算个体的适应度值。这里,选取测试集铜含量预测值与真实值的误差平方和的倒数作为适应度函数,如式(1)所示。其中,针对每一个体在计算 \hat{T} 时,均采用 BP 神经网络法计算,以 92 个训练样本作为神经网络的训练数据,以 29 个测试样本作为神经网络的预测样本。为了避免初始权值和阈值的随机性对适应度函数值的影响,每次计算都用遗传算法对神经网络权值和阈值进行优化。以适应度函数计算结果为依据,以每个个体的适应度值与所有个体适应度之和的比值作为各个体的相对适应度,并以此作为被遗传到下一代的概率,然后采用轮盘赌操作的方式确定个体被选中的次数,最后经过交叉与变异操作,经过多次的迭代计算,输出的末代种群对应的便是筛选出的最具代表性的波段组合。图 3 为适应度变化曲线,由图可知,适应度值的变化曲线在第 22 代收敛,趋向于一条直线,说明遗传算法在第 22 代就找到了最优解。

$$f(X) = \frac{1}{SE} = \frac{1}{\text{sse}(\hat{T} - T)} = \frac{1}{\sum_{i=1}^n (\hat{t}_i - t_i)^2} \quad (1)$$

式(1)中, $\hat{T} = \{\hat{t}_1, \hat{t}_2, \dots, \hat{t}_n\}$ 为样本铜含量的预测值, $T = \{t_1, t_2, \dots, t_n\}$ 为样本铜含量的真实值, n 为本实验测试样本的样本数目。

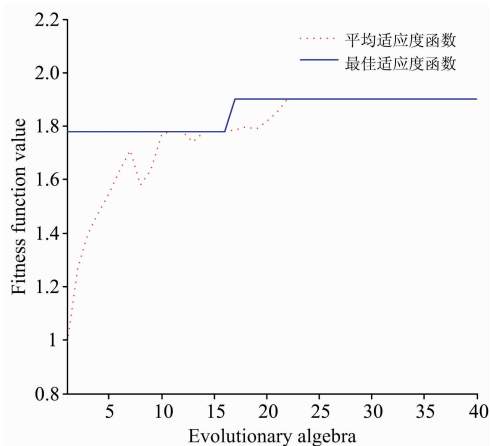


图 3 适应度函数变化曲线

Fig. 3 Fitness function curve

经遗传算法处理后,挑选出了 467 个最佳波段组合,以此为依据,在所有样本的光谱数据中选出对应的 467 个波段反射率作为后续建模的输入数据。

2.2 模型建立与检验

采用含有多个隐藏层的网络模型,其中输入层根据原始数据及经过 PCA, LLE, GA 算法处理后数据的维度分别设置为: 973, 3, 5 和 467, 隐藏层层数 L 可由 $L = \sqrt{M+1} + \alpha$ 计算得出,其中, M 为输入数据的维度, α 取值均设置为 4, 输出层均为 1, 训练函数采用 trainlm, 学习率为 0.000 1, 训练次数为 2 000 次。以 92 个训练样本的数据作为建立神经网络模型的数据。

为验证数据预处理方法对建模的有效性,以测试样本对应的未经预处理与预处理后的数据作为模型的输入数据,输

出预测结果并计算预测值与真实值的平均绝对误差和相对误差,以此作为模型精度的评价指标。表 1 为原始数据及对数数据进行不同预处理后使用 BP 神经网络建模的结果。

表 1 不同预处理方法的反演结果
Table 1 Inversion results of different pretreatment methods

预处理方法	反演绝对误差/%	反演相对误差/%
未处理	0.104	37.2
PCA	0.110	39.4
LLE	0.093	33.3
GA	0.045	16.1

由表 1 可看出,经遗传算法处理后的建模精度最高。如图 4 所示,经遗传算法处理后使用 BP 神经网络预测的预测值与真实值作拟合曲线,预测值与真实值的拟合优度为 0.86,相关系数为 0.92,预测结果较为理想。

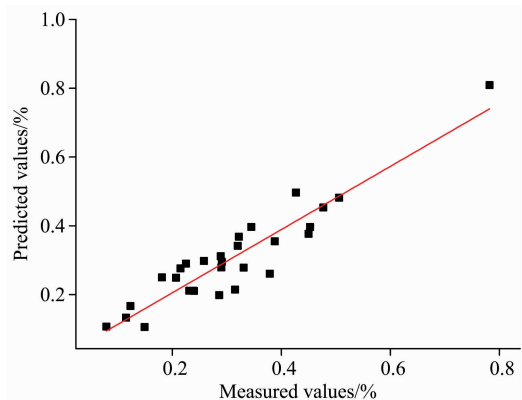


图 4 预测值与实测值对比图

Fig. 4 The comparison of predicted and measured values

由表 1 分析可知,经过预处理后的数据用于建模,反演精度各有不同,表明不同预处理算法对建模结果有不同的影响。综合分析预处理算法原理及反演结果,由于原始数据光谱分辨率高,数据量大,数据间有较高的冗余性,且光谱之间往往存在相关性,将原始数据作为 BP 神经网络模型的输入数据,造成模型精度不高,建模时间长,且易使神经网络出现过拟合现象。主成分分析法作为一种线性降维算法,将原始数据假设为线性结构进行降维,但光谱数据间有大量的非线性结构,这可能是导致主成分分析法处理后的建模效果不佳的主要原因,而局部线性嵌入算法是一种非线性降维算法,该算法假设在局部邻域内是线性的,一定程度上符合光谱数据的非线性结构,因此使用该算法处理后的数据用于建模,模型精度有一定的提高。遗传算法作为一种智能寻优算法,同时与 BP 神经网络结合,将其用作遗传算法中适应度函数关键参数的计算,又以遗传算法对 BP 神经网络的各项权值、阈值进行优化。该算法无需知道先验知识,由算法自动寻找最优波段组合,舍弃敏感性低、对建模可能造成负面效果的波段。结果表明,经该算法处理后的数据用于建模,模

型精度有较大的提高。

对比降维算法与遗传算法对原始数据进行处理后对建模结果的影响, 两类算法均减少了数据量, 简化了模型, 提高了建模速度, 但降维算法改变了原始数据的原始信息及数据结构, 而遗传算法提取了敏感波段, 去除了冗余波段, 保留了数据大部分的原始信息, 并且未改变数据结构, 这可能是遗传算法处理后的数据用于建模, 精度优于两种降维算法的主要原因。

3 结 论

以 121 个乌山斑岩型铜矿的化学分析与可见-光近红外光谱测试数据为数据源, 对测试数据的预处理及定量反演模型的建立进行了深入研究, 得出以下结论:

(1) 用 BP 神经网络对低品位斑岩型铜矿的可见光-近红外光谱数据建立铜品位反演模型具有一定的可行性, 但建模精度偏低。

(2) 用降维算法对数据进行预处理后用于建模, 有效地降低了自变量数量。主成分分析法在建模精度上没有提高, 建模平均绝对误差为 0.110%, 但提高了建模速度。局部线性嵌入算法在提高建模速度的基础上, 同时提高了建模精度, 平均绝对误差为 0.093%。

(3) 对于低品位、光谱特征不明显的斑岩型铜矿, 使用遗传算法与神经网络结合的方法对原始光谱数据进行有效波段提取, 处理后的结果用于建模, 在提高建模速度的同时, 建模精度也有较大提高, 反演平均绝对误差为 0.045%。由于本实验的斑岩型铜矿样本铜含量较低, 平均品位仅为 0.279%, 因此品位反演相对误差达到 16.1%, 这说明该方法处理后所建模型在品位反演方面仍然具有一定的局限性, 但为进一步提高低品位斑岩型铜矿定量反演精度奠定了基础。

本研究对矿山实时, 快速分析矿石品位提供了一定的参考价值, 尤其对低品位斑岩型铜矿品位快速分析具有重要的现实意义。

References

- [1] CHEN Jian-ping, ZHANG Ying, WANG Jiang-xia, et al(陈建平, 张莹, 王江霞, 等). Journal of Geology(地质学刊), 2013, 37(3): 358.
- [2] DENG Hui-juan, JI Gen-yuan, YI Jin-jun, et al(邓会娟, 季根源, 易锦俊). China Mining Magazine(中国矿业), 2016, 25(2): 143.
- [3] Liu J B, Zhang Y, Wang H Y, et al. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2018, 199: 43.
- [4] LÜ Jie, HAO Ning-yan, CUI Xiao-lin(吕杰, 郝宁燕, 崔晓临). Transactions of the Chinese Society of Agricultural Engineering(农业工程学报), 2015, 31(9): 265.
- [5] Shi T Z, Liu H Z, Chen Y Y, et al. Journal of Hazardous Materials, 2016, 308: 243.
- [6] Speta M, Rivard B, Feng J L, et al. AAPG Bulletin, 2015, 99(7): 1245.
- [7] Qu L Q, Han W G, Lin H, et al. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2014, 7(11): 4559.
- [8] GAO Hong-zhi, LU Qi-peng, DING Hai-quan, et al(高洪智, 卢启鹏, 丁海泉, 等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2009, 29(11): 2951.
- [9] An J L, Zhang X R, Zhou H Y, et al. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2018, 11(7): 2513.
- [10] Imani M, Ghassemian H. IEEE Geoscience and Remote Sensing Letters, 2014, 11(8): 1325.
- [11] Liu J B, Han J C, Zhang Y, et al. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2018, 204: 33.
- [12] Liang L, Di L P, Huang T, et al. Remote Sensing, 2018, 10: 10.3390/rs10121940.
- [13] JIANG Tie-cheng(江铁成). Computer Engineering and Sciences(计算机工程与科学), 2016, 38(4): 733.
- [14] Fang Y, Li H, Ma Y, et al. IEEE Geoscience and Remote Sensing Letters, 2014, 11(10): 1712.
- [15] Cui M S, Prasad S, Li W, et al. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2013, 6(3): 1688.

Research on Inversion Model of Low-Grade Porphyry Copper Deposit Based on Visible-Near Infrared Spectroscopy

MAO Ya-chun^{1,2}, DING Rui-bo^{1*}, LIU Shan-jun^{1,2}, BAO Ni-sha^{1,2}

1. College of Resources and Civil Engineering, Northeastern University, Shenyang 110819, China

2. Smart Mine Research Center, Northeastern University, Shenyang 110819, China

Abstract At present, the analysis of copper grades at home and abroad is mainly based on chemical analysis. Due to the disadvantages of high cost, long time and residual pollutants, the chemical analysis method has a serious hysteresis effect on the relative ore blending process, resulting in that the copper content of the tailings is too high, which will inevitably lead to waste of resources. It is an effective way to solve this problem by conducting visible-near-infrared spectroscopy and modeling of porphyry copper deposits. Based on the chemical analysis and spectral test data of 121 Wushan porphyry copper deposits, the visible-near-infrared spectral characteristics of porphyry copper deposits are analyzed. The original spectral data is processed by principal component analysis (PCA) and local linear embedding algorithm (LLE). The reduced dimension is 3 and 5 dimensions respectively. At the same time, the genetic algorithm (GA) is used to select the band of the original spectral data. A total of 467 optimal bands are selected. Then, this paper takes the BP neural network as the modeling method, respectively uses visible-near-infrared spectroscopic data of 92 and 29 porphyry copper deposits as the modeling and testing samples, and establishes a quantitative inversion model of visible-near infrared spectroscopy for porphyry copper deposits. The average absolute error of the grade inversion model based on the original data is only 0.104%. The average absolute error of the grade model based on the model processed by the principal component analysis method, the locally linear embedding algorithm and the genetic algorithm is 0.110%, 0.093% and 0.045% respectively. It can be seen that the grade inversion accuracy of the model based on the data processed by principal component analysis method is poor, the accuracy of grade inversion model based on locally linear embedding algorithm is slightly improved, and the grade inversion accuracy of the model based on the data processed by the genetic algorithm is improved obviously. The research result shows that the grade analysis based on the inversion model of visible-near infrared spectroscopic data of low-grade porphyry copper deposits is feasible, providing an effective method for rapid grade detection of low-grade porphyry copper deposits in China.

Keywords Porphyry copper deposit; Visible-near infrared spectroscopy; Dimensionality reduction algorithm; Genetical algorithm; Inversion model

(Received Jul. 17, 2019; accepted Nov. 30, 2019)

* Corresponding author