激光与光电子学进展

结合灵敏度降维和支持向量回归的土壤元素定量分析方法

李福生[1,2*]，曾小龙[1,2]

[1]电子科技大学自动化工程学院，四川 成都 611731；

[2]电子科技大学长三角研究院，浙江 湖州 313099

摘要 为提高土壤元素定量分析的精度，提出一种结合灵敏度降维与贝叶斯优化算法支持向量回归(BOA-SVR)的土壤元素定量分析方法。利用便携式X射线荧光(XRF)分析仪测量得到土壤的XRF光谱，采用迭代离散小波变换对光谱进行本底扣除，并将计算的各元素净峰面积作为模型输入特征。通过灵敏度分析研究了不同输入特征集合对预测精度的影响，以实现特征降维。将样本分为训练集和测试集，通过均方根误差和决定系数评价模型的预测精度，基于Cu和As元素对比了全特征输入下的BOA-SVR模型、特征降维后的BOA-SVR模型、单参数偏最小二乘法模型的预测结果。实验结果表明，特征降维后的BOA-SVR模型在Cu和As元素预测中都获得最好的预测结果。

关键词 光谱学；X射线荧光光谱；贝叶斯优化算法；支持向量回归；灵敏度分析

中图分类号 TN247 **文献标志码** A **DOI**：10.3788/LOP213241

Quantitative Analysis Method of Soil Elements Combining Sensitivity Dimensionality Reduction and Support Vector Regression

Li Fusheng[1,2*], Zeng Xiaolong[1,2]

[1]*School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu* 611731, *Sichuan, China*;

[2]*Yangtze Delta Region Institute, University of Electronic Science and Technology of China, Huzhou* 313099, *Zhejiang, China*

Abstract This study proposes a quantitative analysis method combining sensitivity dimensionality reduction and Bayesian optimization algorithm support vector regression (BOA-SVR) to improve quantitative analysis accuracy of soil elements. The X-ray fluorescence (XRF) spectrum of the soil is obtained using a portable XRF analyzer, and the background is subtracted by iterative discrete wavelet transform. Furthermore, the calculated net peak area of each element is used as the model input feature. The influence of different input feature sets on the prediction accuracy is studied using sensitivity analysis to achieve feature dimensionality reduction. The samples are divided into training and test sets, and prediction accuracy of the model is evaluated using the root mean square error and coefficient of determination. Based on Cu and As elements, the prediction results of the BOA-SVR model under full feature input, the BOA-SVR model after feature dimension reduction, and the single-parameter partial least squares model are compared. The experimental results show that BOA-SVR model after feature dimension reduction achieves the best prediction result in both Cu and As elements.

Key words spectroscopy; X-ray fluorescence spectrum; Bayesian optimization algorithm; support vector regression; sensitivity analysis

1 引 言

土壤作为一个复杂而独立的生态系统，可以与周围环境进行物质和能量交换，其质量与植物、动物和人类的健康密切相关。近年来，随着人口的不断扩大和社会经济的发展，土地的开发利用程度不断提高，土壤

收稿日期：2021-12-15；**修回日期**：2022-01-26；**录用日期**：2022-02-28；**网络首发日期**：2022-03-10

基金项目：国家自然科学基金(62075028)

通信作者：*lifusheng@uestc.edu.cn

退化问题也日益严重，如土壤肥力下降、土壤污染加剧[1]。

土壤质量与土壤中各种元素的含量密切相关，除了重金属元素含量对土壤污染程度的决定性影响[2-3]，反映土壤肥力状况的有机质（OM）、全氮（TN）、速效磷（AP）、速效钾（AK）土壤盐分等特征与土壤中的Ca、Cu、Fe、Zn、Mn元素息息相关[4]，反映土壤酸碱度的pH值与Mn、P、Hg、Cd等元素密切相关[5-6]。因此，一种有效的土壤元素含量测量方法对判定土壤质量、合理利用土壤资源等具有重要意义。土壤中元素含量的常规测量分析方法有分光光度法、电感耦合等离子体原子发射光谱法和原子吸收光谱法等[7-8]，这些方法都比较成熟，且精度高、结果可靠，但也存在较多问题，如对样本的预处理过程繁琐，且在处理过程中需要使用大量化学处理方法，容易对环境造成二次污染。X射线荧光（XRF）光谱检测技术是一种元素含量分析技术，具有成本低、速度快、元素范围广等优势，成为土壤元素含量分析的主要方法之一，目前已被广泛应用于物理、化学、生物、环境、工业生产等领域[9-10]。

在基于XRF光谱的元素定量分析中，对于土壤中元素的含量计算，主要通过测量元素对应的特征峰强度建立校准曲线，如通过偏最小二乘（PLS）法或常规的数学分析方法对谱线强度与元素含量的关系进行拟合分析。但元素间的基体效应、光电吸收增强效应、数据自身的本底噪声等因素会使元素含量的XRF测量是一个非线性过程，难以准确预测，而单变量线性校准方法PLS难以准确拟合分析信号和研究参数之间的非线性关系。针对这些情况，人们提出了一些非线性校准方法，如支持向量回归（SVR）、反向传播（BP）、AdaBoost算法[11-12]，以提供更好的模型调整能力和预测结果。此外，在XRF定量分析中一般无法获取大量且全面的标定样本进行测量，使很多基于统计的神经网络方法难以取得很好的效果。SVR算法作为一种多元非线性回归算法，具有很高的泛化性能，在小样本集中也能提供良好的模型。

本文研究了XRF技术结合SVR算法定量分析土壤中各元素含量的可行性。首先，基于贝叶斯优化算法（BOA）对SVR模型进行优化，解决了SVR对超参数的敏感性和依赖性问题。然后，针对XRF光谱数据，采用迭代离散小波变换（IDWT）对样本进行预处理，计算样本中各元素的净峰面积并将其作为模型输入提高预测精度。同时，研究了不同输入特征对模型预测性能的影响，提出了基于灵敏度分析的特征降维方法，基于最优的输入变量和模型参数构建定量分析模型。最后，将模型应用于土壤元素的定量分析中，并与单参数PLS算法的预测结果进行了对比。

2 实验原理

2.1 支持向量回归算法

基于XRF的土壤元素定量分析中，谱图的测量误差、元素间的基体效应等会导致元素与特征峰强度间呈现非线性关系，寻找大量且全面的土壤标定样本比较困难，常常会出现小样本集情况。SVR算法作为一种多元非线性回归算法，具有较好的泛化性能，即使在小样本集情况下也能提供良好的模型，因此，实验采用SVR算法作为分析模型。SVR算法是在支持向量机（SVM）基础上发展的算法，其目的是通过寻找一个结构风险最小化的映射函数拟合输入输出间的非线性关系。SVR算法通过将低维空间中难以解决的非线性问题映射到高维空间中，将非线性回归转换为高维特征空间的线性问题，从而简化问题的求解过程[13]。

2.2 贝叶斯优化算法

SVR算法的预测效果非常依赖模型的超参数，如惩罚系数C、核函数类型γ、不敏感损失函数ε，且这些超参数必须由用户进行充分调整，过程比较复杂。为了获得SVR模型的全局最优性能，采用BOA搜索SVR模型最合适的超参数，即最佳的土壤元素含量预测模型。BOA是一种寻找黑盒函数极值的强大技术，可以很好地解决目标函数评估代价昂贵的问题。典型的BOA根据已知观测值拟合高斯过程模型，然后利用采集函数得到目标函数极值的后验位置。其中，采集函数是BOA的关键参数，负责下一个潜在最大值点的选择，同时决定了算法的勘探与开发性能。目前，置信上限期望优化（UCB）在寻找多模态黑匣子函数的全局最优所需的函数评估数量上也被验证是有效的[14]，因此，实验中用UCB算法作为采集函数。

将SVR模型5倍交叉验证结果的均方根误差作为BOA的目标损失函数。BOA的输入变量是SVR的核函数类型γ、惩罚系数C、核函数的Gamma系数。核函数类型有线性核函数、多项式核函数、径向基核函数3种，惩罚系数和Gamma系数的变化范围均设置为[0.01,100]。BOA的输出变量是SVR模型5倍交叉验证的准确率[14]。

2.3 光谱预处理方法与元素净峰面积求解

SVR算法作为一种定量算法，相比其他回归算法，在样本数量较少情况下具有较好的预测精度。为了获取更好的预测性能，需要对样本质量进行控制，且SVR算法本身是基于SVM生成的估计函数，这也体现出其对训练样本质量的敏感性。实际测量中，大多数测量信号都含有噪声等干扰，这对单元分析有很大的影响。XRF光谱中本底噪声是最严重的干扰，光谱背景往往使净峰面积估计结果过大，峰值位置估计结果发生偏移，因此在定量分析前必须扣除本底。已有研究表明，离散小波变换（DWT）在土壤样品光谱信号本底扣除

和去噪方面表现出良好的效果[15]。因此，采用IDWT对55个样本的光谱进行处理。IDWT是一种扣除光谱本底十分有效的DWT算法，具体流程如下。

1）确定小波基类型和小波变换的分解层数。

2）用$f_m[i]$表示经过$m-1$次DWT迭代处理后得到的光谱。其中：m为大于等于1的整数；i为光谱通道；$f_1[i]$为原始的XRF光谱数据。将m初始化为1，同时指定最小误差ε作为迭代停止的条件。

3）通过DWT算法将第i次迭代后的光谱$f_m[i]$分解为多个频段的信号分量。

4）选择第1层低频分量$u_1[i]$作为$e_m[i]$，即$e_m[i]=u_1[i]$。其中，$e_m[i]$为第m次迭代后算法估计的本底。

5）如果$m=1$，跳至步骤6）；否则，计算$\left|e_m[i]-e_{m-1}[i]\right|_{\max}=e^m$。将$e^m$和$\varepsilon$进行对比，$t$的初始化值为0。若$e^m<\varepsilon$，表明相邻两次所估计的本底足够一致，令$t=t+1$；否则，重置$t=0$。

6）如果$t<3$，跳至步骤7）；如果$t\geqslant 3$，表明连续3次经过DWT分解估计的本底足够一致并满足精度要求，即本底噪声已经收敛，取最终一次DWT分解估计的本底$e_m[i]$作为最终估计的本底，从原光谱$f_1[i]$中减去$e_m[i]$即可实现本底扣除。

7）对比$f_m[i]$和$e_m[i]$的所有通道值，将两光谱中各通道的最小值替换为$f_m[i]$。

实验选择sym4小波基，设置的分解层数为9，循环迭代后可以准确扣除背景，利用IDWT算法对土壤光谱数据的处理结果如图1所示。图1(a)为元素谱图和IDWT算法最终估计的本底。其中，x轴为能量坐标系，其坐标和元素的特征X射线一一对应。可以发现，估计的本底与实际本底的拟合度较好。原始谱图和本底扣除校正后的谱图如图1(b)所示。

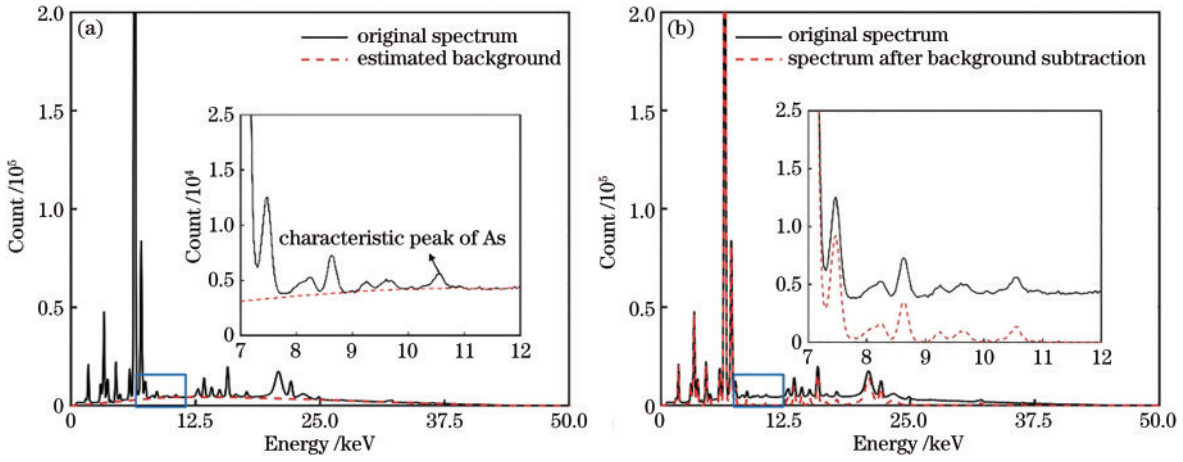

图1　本底扣除前后的光谱。(a)原始光谱与估计的本底；(b)原始光谱与校正后的光谱

Fig. 1　Spectra before and after background subtraction. (a) Original spectrum and estimated background; (b) original spectrum and corrected spectrum

图1中在能量为10.53 keV位置的峰表示As元素的特征峰。可以发现，As元素的含量很少，导致峰强度较低，但其本底非常大，严重影响对该元素特征峰强度的计算。因此，小波预处理后根据不同元素的峰值通道计算净峰面积，将每个元素的净峰面积作为输入特征以进一步提高定量分析的精度。计算样本中29个常见元素的净峰面积作为输入特征，以As元素为例，经过小波预处理前后As元素的实际含量与As元素净峰面积的关系及拟合曲线如图2所示。可以发现，使用小波变换后决定系数R^2由0.085大大提高到0.778，实际含量与输入特征的线性关系明显更强。

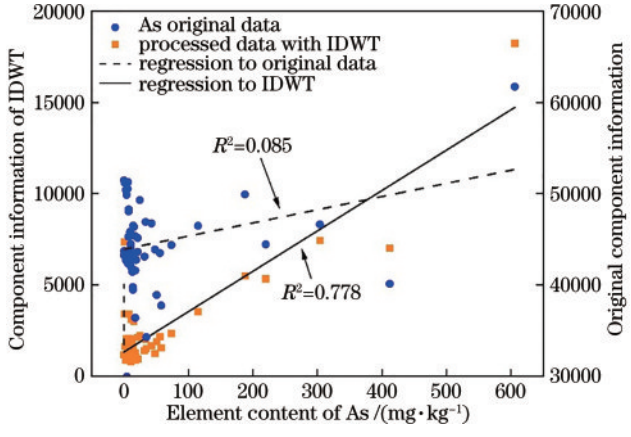

图2　本底扣除后的预处理效果

Fig. 2　Preprocessing effect after background subtraction

2.4　基于灵敏度分析的特征降维方法

在XRF元素含量定量分析中，元素间基体效应、仪器自身噪声等会导致元素净峰面积的估计误差。其中，谱线干扰是定量分析中的主要误差源之一，当元素的特征X射线波长几乎相等时，X射线强度测量将受到干扰。元素的定量分析主要受重叠峰和逃逸峰的影响。同时，变量太多也不利于SVR模型的建立，严重的变量间共线关系会影响模型的准确性和稳定性。为

了进一步提高模型的定量分析能力，有必要去除XRF光谱中无用的特征，并筛选出与被测元素定量分析相关的特征进行建模。采用灵敏度分析方法对模型进行特征降维，具体步骤如下。

1）利用贝叶斯优化算法搜索得到所有29个特征作为输入的最优模型超参数。

2）基于最优超参数和部分样本训练得到SVR模型。

3）取用步骤2）中的所有样本，针对第i个输入特征c，将样本中特征c的数据分别增大和减少10%后获得样本集e_1和e_2。

4）利用SVR模型分别预测e_1和e_2，得到预测结果集r_1和r_2，以及特征c的灵敏度$S_i=\mathrm{abs}(r_1-r_2)$。其中，abs(•)为求绝对值函数。

5）对所有输入特征，重复步骤3）和步骤4），最终获得所有特征的灵敏度系数$S_i(1\leqslant i\leqslant 29)$，对所有的灵敏度系数按从大到小排序得到$S_i^*(1\leqslant i\leqslant 29)$，最后选择最高的$k$个特征作为降维后的模型输入，从而提高模型精度。其中，k满足

$$\sum_{i=1}^{k}S_i^*\geqslant\sum_{i=1}^{29}S_i^*\times 85\%。\tag{1}$$

2.5　基于BOA-SVR的XRF定量分析策略

基于灵敏度分析和BOA-SVR提出了一种新的XRF元素定量分析策略。针对光谱样本数据，先用迭代小波扣除本底并计算各元素净峰面积，然后基于灵敏度分析方法计算得到各特征对被测元素的灵敏度，最后从大到小选择灵敏度总和占比高于90%的元素作为输入特征建立模型。该定量分析方法的具体流程如图3所示。

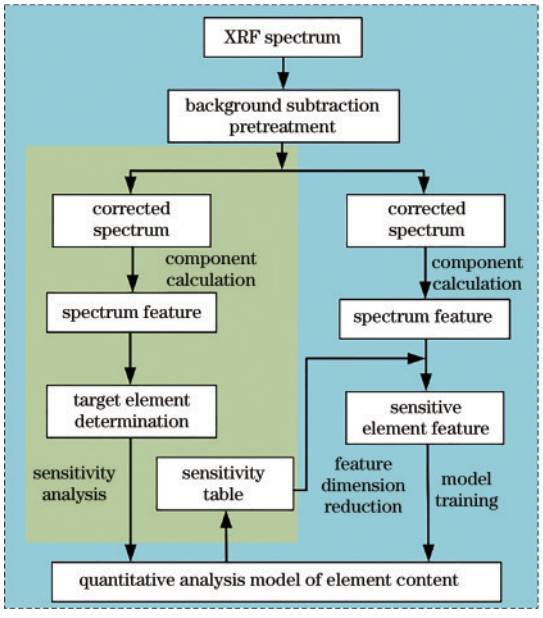

图3　基于BOA-SVR的定量分析方法流程图

Fig. 3　Flow chart of quantitative analysis method based on BOA-SVR

3　实验部分

3.1　仪器与样本

实验使用的设备是由泰克松德公司生产制造的手持式ED-XRF光谱仪，型号为TS-XH4000-SOIL，X射线管的工作电压为45 kV，工作电流为25 μA。采用55个国标样品作为土壤标准样品，样本中每种待测元素的质量分数都具有足够宽的内容范围和适当的内容梯度。

3.2　样品制备与测量

首先，让土壤样品在空气中干燥，然后反复研磨筛选后在120 ℃下干燥24 h。将99%以上的样品用高铝陶瓷球磨机磨至出料粒度为0.074 mm，然后放入样品杯中，手持式光谱仪和压制好的样本如图4所示。最后，用光谱仪对样品进行扫描和测试。

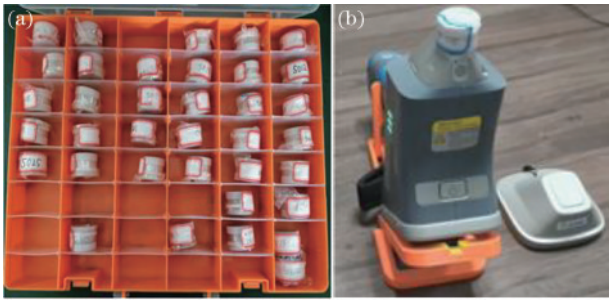

图4　样本和XRF光谱仪的实物图。(a)样本；(b)XRF光谱仪

Fig. 4　Physical image of the sample and XRF spectrometer. (a) Sample; (b) XRF spectrometer

3.3　基于灵敏度分析的特征降维方法

为了筛选出与被测元素定量分析相关的特征建立模型，采用灵敏度分析方法进行特征降维。以As元素为例，灵敏度分析后得到的测试结果如图5所示。可以发现，29种输入元素的净峰面积中Fe、Co、Ti、As、Sb、Ca、Pb对As元素定量分析的影响最大，该结果在物理分析中也能得到一定验证。其中，As元素的分析

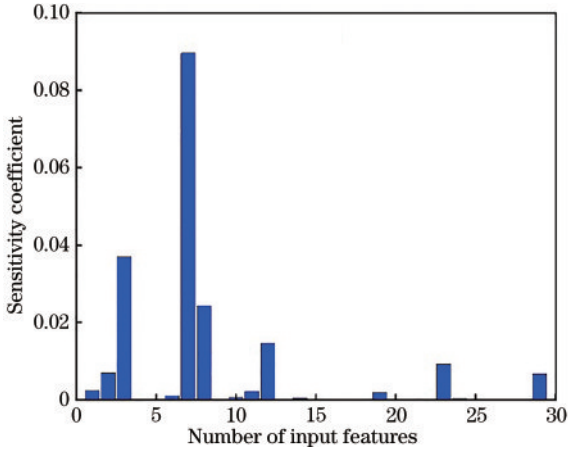

图5　As元素的灵敏度分析结果

Fig. 5　Sensitivity analysis result of the As element

受到As元素自身净峰面积的影响，但影响程度不高，原因是As在土壤中是相对微量的元素。同时，由于As的Kα峰和Pb的Lα峰重叠，As也会受到Pb元素的影响，在As中加入Pb可以进一步提高定量分析精度。而Fe元素是土壤元素中成分占比最大的元素，因此，Fe的净峰面积也是分析中的一个重要输入特征。

为了验证特征降维的有效性，各特征灵敏度排序后依次取前i($1\leqslant i\leqslant 29$)维最高灵敏度的特征，将1个到完整的29个特征分别作为输入参数用于模型训练，最后利用基于BOA-SVR的留一交叉验证法对比不同特征维度下模型的精度，并以留一交叉验证决定系数R_{CV}^2和留一交叉验证均方根误差R_{CV}^{MSE}作为评价指标，结果如图6所示。可以发现：当特征维度较小时($i<5$)，即使选择灵敏度最高的维度也无法较好拟合出被测元素，R_{CV}^{MSE}值非常高；随着输入特征变多，模型精度逐渐提高，特征维度为7时模型精度最高，此时，R_{CV}^{MSE}和R_{CV}^2分别达到最小值和最大值，输入特征刚好包含了灵敏度最大的7个特征，即Fe、Co、Ti、As、Sb、Ca、Pb；随着特征维度的进一步加大，很多与As元素基本不相关的特征引入，反而影响了模型的准确性和稳定性，在维度为20时R_{CV}^2急剧下降。原因是BOA会基于5倍交叉验证的误差为每个维度的SVR模型寻找最优超参数，引入一些随机性误差，且R_{CV}^2只能表示真实与预测结果的线性关系的强弱，不能完全代表误差的大小。该实验结果验证了灵敏度分析的有效性和特征降维方法的必要性。

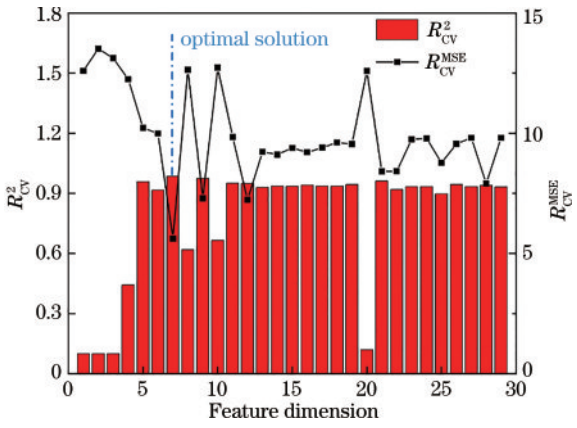

图6　不同特征维度下模型的预测结果

Fig. 6　Prediction results of the model under different feature dimensions

4　实验结果验证与分析

选择重金属元素Cu和微量重金属元素As作为待测元素对所提算法进行验证。首先，对Cu元进行定量分析验证，验证时将实验样品分为训练集和测试集两个集合，分别用于外部验证和内部验证。为了评价预测模型的性能，采用训练集均方根误差R_C^{MSE}、训练集决定系数R_C^2、测试集均方根误差R_P^{MSE}和测试集决定系数R_P^2作为评价指标。然后，基于灵敏度分析得出Cu元素主要受到Fe、Co、Ni、Cu等净峰面积的影响的结论，选择最优输入特征为这4种元素。用最优输入特征和全部特征作为输入，基于BOA找到最优模型参数，建立预测土壤样品Cu元素含量的SVR定量预测模型。同时以全部特征作为输入建立了单参数PLS模型，通过5倍交叉验证(CV)选择单参数PLS模型的最优主成分个数为9。基于校准集数据分别建立了三种模型，利用这些模型对13个测试集和42个训练集数据中的Cu元素含量进行预测，结果如图7所示。

三种模型在13个测试集样本的详细预测结果和相对误差如表1所示，三种模型的整体性能参数如表2所示。其中，SVR*表示经过特征降维的SVR模型，R_C^{MSE}和R_C^2表示直接对训练集样本进行预测后得到的均方根误差和决定系数，R_P^{MSE}和R_P^2表示对测试集样本预测后得到的均方根误差和决定系数。可以发现：对训练集数据进行直接预测时，采用全部特征作为输入的SVR模型取得了最好的效果，其预测结果和原数据几乎一致($R_C^2=0.9988$，$R_C^{MSE}=6.9356$)；基于4个高灵敏度特征作为输入的SVR*模型精度稍次于上一个模型($R_C^2=0.9970$，$R_C^{MSE}=11.0334$)；PLS模型在训练集预测中最差($R_C^2=0.9856$，$R_C^{MSE}=24.1391$)。

对于测试集数据，采用全部特征作为输入的SVR模型获得了非常差的结果($R_P^2=0.9146$，$R_P^{MSE}=73.8296$)，从图7可以发现：Cu的真实含量和预测含量的拟合曲线决定系数低($R_P^2=0.9146$)，表明预测过程的随机性较大；Cu的真实含量和预测含量之间误差大($R_P^{MSE}=73.8296$)，表明预测精度差。虽然全部输入特征为SVR提供了更多的信息，使模型能尽可能地拟合训练集数据，但较多特征与Cu含量的预测不相关甚至引入了噪声信息，使SVR模型在预测测试集时的效果较差。因此需要特征降维筛选出与Cu含量相关的特征，基于4个高灵敏度特征的SVR在预测测试集时获得了良好的效果($R_P^2=0.9918$，$R_P^{MSE}=22.8803$)，拟合曲线的决定系数很高($R_P^2=0.9918$)，这表明Cu的预测含量和真实含量的一致性较好，同时平方根误差系数($R_P^{MSE}=22.8803$)远远低于PLS和全特征SVR模型，表明模型预测的含量与实际含量基本一致。PLS在测试集的预测中结果不是很好($R_P^2=0.9315$，$R_P^{MSE}=66.1133$)，虽然决定系数一致性较好，但数据整体表现出偏差，尤其在元素含量较高的数据上，如Cu元素的最高含量为916，但预测结果为1062，偏差非常大。

对As元素做相同的定量分析实验验证，根据灵敏度分析结果得到Fe、Co、Ti、As、Sb、Ca、Pb与As元素测量最相关的7种元素特征。然后基于校准集数据分别建立三种模型，利用这些模型对测试集和训练集数据中的Cu元素含量进行预测，结果如图8所示，三种模型的整体参数对比如表3所示。可以发现，As元素

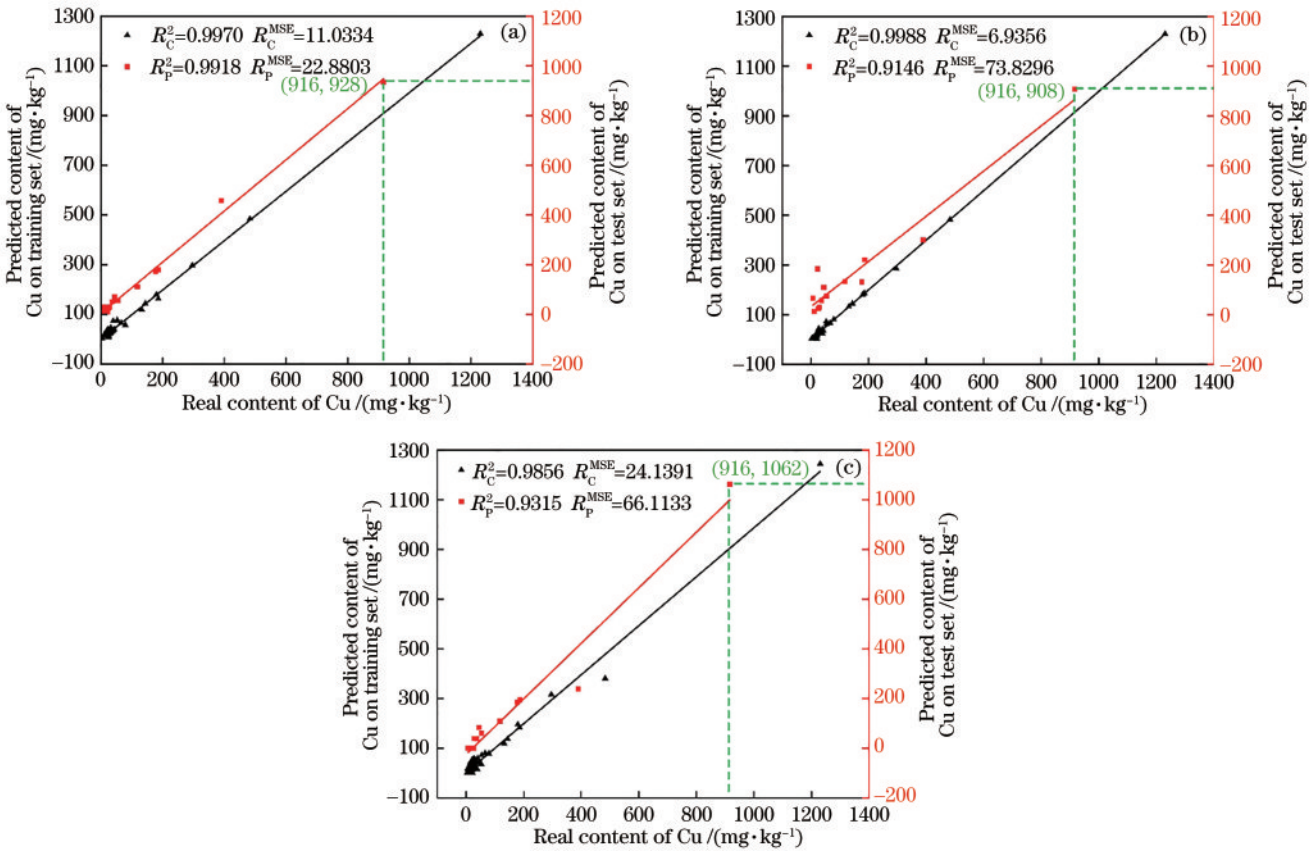

图 7　Cu元素的预测结果。(a)经过特征降维的SVR模型;(b)全部特征作为输入的SVR模型;(c) PLS模型

Fig. 7　Prediction results of Cu element. (a) SVR model with feature dimension reduction; (b) SVR model with all features as inputs; (c) PLS model

表 1　Cu含量预测时三种模型在验证集上的预测结果

Table 1　Prediction results of three models on verification set in Cu element verification

No.	Reference value	Predictive value			Relative error		
		SVR*	SVR	PLS	SVR*	SVR	PLS
2	25.7	23.97	23.37	0.00	0.0672	0.0906	1.0000
5	916.0	932.84	908.24	1062.30	0.0184	0.0085	0.1597
7	7.2	9.84	66.27	0.00	0.3671	8.2047	1.0000
10	187.0	179.84	220.50	195.00	0.0383	0.1791	0.0428
12	54.2	54.88	74.23	60.07	0.0125	0.3696	0.1082
21	22.6	10.60	184.53	0.00	0.5308	7.1650	1.0000
24	118.0	112.03	133.18	108.47	0.0506	0.1286	0.0807
27	45.0	70.77	109.63	82.73	0.5727	1.4362	0.8384
32	177.0	170.59	131.64	184.14	0.0362	0.2563	0.0403
33	37.0	50.48	56.65	37.60	0.3642	0.5312	0.0162
48	29.0	29.44	29.54	39.15	0.0150	0.0187	0.3498
52	390.0	457.35	300.90	237.89	0.1727	0.2285	0.3900
53	11.4	19.69	12.83	0.00	0.7270	0.1259	1.0000

的测量效果比Cu元素的测量效果整体较差。原因是As元素在土壤中的含量较少,导致仪器自身和测量过程中的噪声会对As元素的测量带来较大的影响。与Cu元素的预测结果相同,对训练集数据进行预测时,采用全部特征作为输入的SVR模型取得了最好的效果,其预测结果和真实数据几乎一致(R_C^2=0.9996,R_C^{MSE}= 0.3038),但是模型的训练出现了过拟合的问题,导致在对测试集数据预测时表现非常差(R_P^2=0.7534,R_P^{MSE}=16.5271),模型泛化能力很弱。而经过灵敏度降维处理的SVR模型在三种模型中的性能最

表 2　Cu 含量预测时三种模型的预测结果

Table 2　Cu element prediction results obtained by three models

Model	R_C^{MSE}	R_C^2	R_P^{MSE}	R_P^2
SVR*	11.0334	0.9970	22.8803	0.9918
SVR	6.9356	0.9988	73.8296	0.9146
PLS	24.1319	0.9856	66.1133	0.9315

表 3　As 含量预测时三种模型的预测结果对比

Table 3　As element prediction results obtained by three models

Model	R_C^{MSE}	R_C^2	R_P^{MSE}	R_P^2
SVR*	1.1271	0.9863	11.6868	0.9526
SVR	0.3038	0.9996	16.5271	0.7534
PLS	17.0948	0.4192	37.5909	0.4899

好，在训练集和测试集的预测中都获得了相对较好的预测结果（R_C^2=0.9863，R_C^{MSE}=1.1271，R_P^2= 0.9526，R_P^{MSE}=11.6868）。这表明将高灵敏度特征作为输入变量构建 BOA-SVR 定量模型是定量分析土壤中的元素含量的一种可行方法。

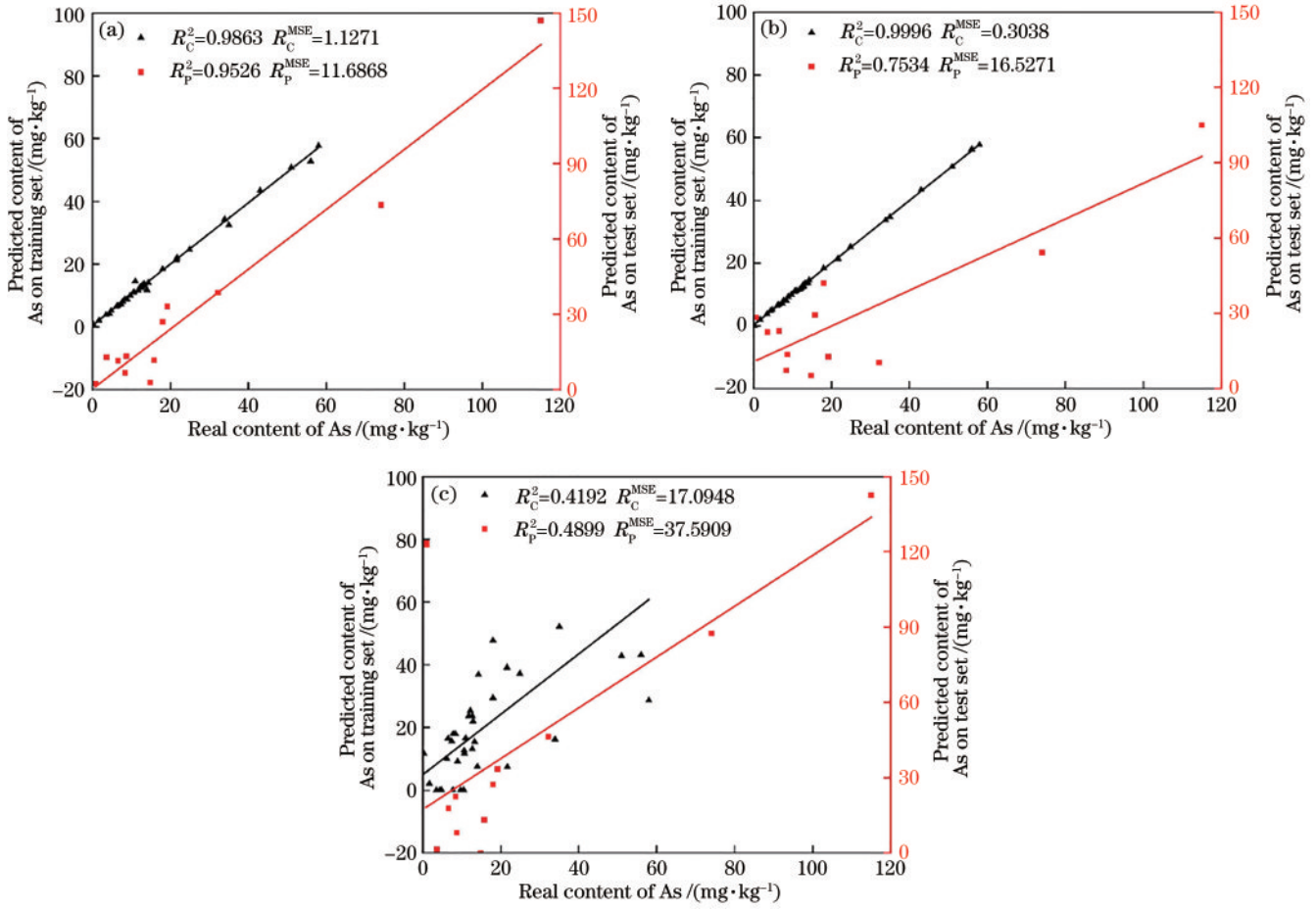

图 8　As 元素的预测结果。(a)经过特征降维的 SVR 模型；(b)全部特征作为输入的 SVR 模型；(c) PLS 模型

Fig. 8　Prediction results of As element. (a) SVR model with feature dimension reduction; (b) SVR model with all features as inputs; (c) PLS model

5　结　　论

主要验证 XRF 技术结合 SVR 算法定量分析土壤中各元素含量的可行性。针对 XRF 光谱样本数据，采用 IDWT 对样本进行预处理，并将计算的样本中各元素的净峰面积作为模型输入提高预测精度。针对被测元素，基于灵敏度分析方法获得与被测元素测量相关的特征，实现特征降维。基于最优输入变量和 BOA-SVR 构建最优定量分析模型，将模型应用于对土壤中 Cu 元素和 As 元素的定量分析，分析实验中，采用高灵敏度的 Fe、Co、Ni、Cu 4 个特征作为输入和以 Fe、Co、Ti、As、Sb、Ca、Pb 7 个特征作为输入的 SVR 模型分别在定量分析中获得了最好的预测结果。结果表明，基于灵敏度分析的特征降维方法可以剔除与被测元素无关的特征和噪声数据，提高模型精度。使用 55 个土壤样品用于定量分析，以 Cu 元素的分析为例，样本中 Cu 元素的质量分数主要集中在 200 mg/kg 以下且分布不均匀，对比算法性能时，采用外部验证或交叉验证会导致验证结果不稳定。此外，SVR 模型训练预测时很难学习到高含量样本的信息，导致预测高含量样本时模

型的效果较差，如出现欠拟合的问题，可使用蒙特卡罗算法生成模拟光谱样本数据解决该问题。

参 考 文 献

[1] Tepanosyan G, Sahakyan L, Belyaeva O, et al. Continuous impact of mining activities on soil heavy metals levels and human health[J]. Science of the Total Environment, 2018, 639: 900-909.

[2] Zhang X Y, Zhong T Y, Liu L, et al. Impact of soil heavy metal pollution on food safety in China[J]. PLoS One, 2015, 10(8): e0135182.

[3] Huang J H, Guo S T, Zeng G M, et al. A new exploration of health risk assessment quantification from sources of soil heavy metals under different land use[J]. Environmental Pollution, 2018, 243: 49-58.

[4] 史进, 李文胜, 张俊苗. 两种果园土壤质量综合评价及生物量与土壤元素的关系[J]. 新疆农业科学, 2016, 53(6): 1081-1090.
Shi J, Li W S, Zhang J M. Comprehensive evaluation of the soil quality and the relationship between biomass and soil elements in two orchards[J]. Xinjiang Agricultural Sciences, 2016, 53(6): 1081-1090.

[5] Lu C P, Lv G, Shi C Y, et al. Quantitative analysis of pH value in soil using laser-induced breakdown spectroscopy coupled with a multivariate regression method[J]. Applied Optics, 2020, 59(28): 8582.

[6] Zhao M J, Yan C H, Feng Y Z, et al. A novel strategy for quantitative analysis of soil pH via laser-induced breakdown spectroscopy coupled with random forest[J]. Plasma Science and Technology, 2020, 22(7): 074003.

[7] Shard A G, Wright L, Minelli C. Robust and accurate measurements of gold nanoparticle concentrations using UV-visible spectrophotometry[J]. Biointerphases, 2018, 13(6): 061002.

[8] Tekin Z, Unutkan T, Erulaş F, et al. A green, accurate and sensitive analytical method based on vortex assisted deep eutectic solvent-liquid phase microextraction for the determination of cobalt by slotted quartz tube flame atomic absorption spectrometry[J]. Food Chemistry, 2020, 310: 125825.

[9] Stankey J A, Akbulut C, Romero J E, et al. Evaluation of X-ray fluorescence spectroscopy as a method for the rapid and direct determination of sodium in cheese[J]. Journal of Dairy Science, 2015, 98(8): 5040-5051.

[10] Zhou S B, Yuan Z X, Cheng Q M, et al. Quantitative analysis of iron and silicon concentrations in iron ore concentrate using portable X-ray fluorescence (XRF)[J]. Applied Spectroscopy, 2020, 74(1): 55-62.

[11] 宋海声, 陈召, 徐大诚, 等. GA-BP 神经网络结合 EDXRF 技术实现对中低合金钢中 Cr、Mn 和 Ni 元素含量的预测[J]. 激光与光电子学进展, 2022, 59(12): 544-550.
Song H S, Chen Z, Xu D C, et al. Prediction of Cr, Mn, and Ni in medium and low alloy steels by GA-BP neural network combined with EDXRF technology[J]. Laser & Optoelectronics Progress, 2022, 59(12): 544-550.

[12] Li F S, Yang W Q, Ma Q, et al. X-ray fluorescence spectroscopic analysis of trace elements in soil with an adaboost back propagation neural network and multivariate -partial least squares regression[J]. Measurement Science and Technology, 2021, 32(10): 105501.

[13] 尚栋, 孙兰香, 齐立峰, 等. 基于循环变量筛选非线性偏最小二乘的 LIBS 铁矿浆定量分析[J]. 中国激光, 2021, 48(21): 2111001.
Shang D, Sun L X, Qi L F, et al. Quantitative analysis of laser-induced breakdown spectroscopy iron ore slurry based on cyclic variable filtering and nonlinear partial least squares[J]. Chinese Journal of Lasers, 2021, 48(21): 2111001.

[14] Schmidt-Hieber J. Nonparametric regression using deep neural networks with ReLU activation function[J]. The Annals of Statistics, 2020, 48(4): 1875-1897.

[15] Li F, Lu A X, Wang J H. Modeling of chromium, copper, zinc, arsenic and lead using portable X-ray fluorescence spectrometer based on discrete wavelet transform[J]. International Journal of Environmental Research and Public Health, 2017, 14(10): 1163.