

PSO-LSSVM 对 LIBS 定量分析精度的提高

林晓梅¹, 王晓檬¹, 黄玉涛^{1*}, 林京君^{2*}

1. 长春工业大学电气与电子工程学院, 吉林 长春 130012
2. 长春工业大学机电工程学院, 吉林 长春 130012

摘要 针对土壤定量分析受基体效应影响大, LIBS 定量分析精度不佳等问题, 采用粒子群算法对 LSSVM 进行优化, 提高模型的精确度。选取 Pb I 405.78 nm 和 Cr I 425.44 nm 作为分析谱线进行分析。采集十二个不同浓度样品的特征光谱, 每个浓度样品在不同点采集 20 组数据, 将其中 17 组数据设为训练集, 3 组数据设为预测集, 用 LSSVM 和 PSO-LSSVM 两种方法建立定标模型。对比两种模型的拟合相关系数 (R^2)、训练集均方根误差 (RMSEC) 和预测集均方根误差 (RMSEP)。由于自吸收效应的影响, 随着浓度的增加, 预测值逐渐低于实际值, LSSVM 定标模型的拟合程度较低, 无法达到实验要求, 模型性能有待提高。利用粒子群算法对 LSSVM 的模型参数惩罚系数和核函数参数进行优化, 得到最佳的参数组合, Pb 元素为 (8 096.8, 138.865 7), Cr 元素为 (4 908.6, 393.563 5), 用最佳的参数组合构建 LSSVM 的定标模型。相比于 LSSVM, PSO-LSSVM 定标模型的精确度更高, Pb 和 Cr 元素的 R^2 提高到了 0.982 8 和 0.985 0, 拟合效果明显提升。Pb 和 Cr 元素的训练集均方根误差由 0.026 0 Wt% 和 0.027 2 Wt% 下降到 0.022 4 Wt% 和 0.019 1 Wt%, 预测集均方根误差由 0.101 8 Wt% 和 0.078 8 Wt% 下降到 0.045 8 Wt% 和 0.042 0 Wt%, 模型的稳定性进一步提高。说明 PSO-LSSVM 算法能够更好地降低土壤基体效应和自吸收效应带来的影响, 提高分析结果的精确度与稳定性。

关键词 激光诱导等离子体技术; 粒子群优化; 最小二乘支持向量机; 定量分析

中图分类号: O433.4 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2021)11-3583-05

引言

土壤是发展经济的重要物质基础, 但我国土壤受重金属污染程度日益严重。因此, 如何提高土壤重金属定量分析的稳定性和精确度十分重要。随着激光技术的发展, LIBS 技术逐渐被应用到土壤的定量分析中^[1]。

土壤定量分析受基体效应影响较大, 因此应用 LIBS 技术定量分析时, 通常需要结合一些分析方法来提高 LIBS 的分析精度^[2]。Villas-Boas 等^[3]将偏最小二乘回归方法与 LIBS 技术结合分析土壤中金属元素, 拟合系数均高于 0.85。汪萍等^[4]用 LIBS 技术和三元二阶非线性回归方法对土壤中 Mn 元素进行检测, 测量浓度与参考浓度的相关系数为 0.97。De 等^[5]用 LIBS 技术和内标法结合检测土壤中的 Ca, 拟合系数提高到 0.989。孟德硕等^[6]将 LIBS 技术与 ANN 结

合检测土壤中的 Cu, 测量误差在 10% 以下。邹孝恒等^[7]采用 LIBS 技术与遗传算法和偏最小二乘法相结合, 提高了土壤中金属元素的检测精确度。

近年来, 诸如卷积神经网络 (CNN) 等人工智能算法被广泛应用于检测土壤中的元素^[8]。但是, 卷积神经网络存在平移不变性、池化层和过度拟合等缺点; 最小二乘支持向量机 (LSSVM) 作为新一代机器学习算法, 能够避免神经网络过拟合和支持向量机训练耗时长的问题^[9]。然而, LSSVM 的参数搜索是采用网格搜索法选择模型参数并进行交叉验证, 搜索精度低和效率慢都是此算法的缺陷^[10]。LSSVM 建模过程中, 对预测结果影响较大的参数为惩罚系数和核函数参数 g ^[11]。LSSVM 模型中的参数一般是根据经验来设定的, 盲目性大和效率低都是需要解决的问题^[12]。为了克服 LSSVM 算法存在的缺陷, 需要采用其他算法优化模型参数, 得到最优的参数组合 (γ, g)。常用的优化算法一般操作步骤繁琐、

收稿日期: 2020-10-29, 修订日期: 2021-02-11

基金项目: 国家重大科学仪器开发专项(2014YQ12035104), 吉林省科技厅项目(20180414017GH, 20200403008SF), 吉林省发展改革委项目(2018C034-3)资助

作者简介: 林晓梅, 女, 1965年生, 长春工业大学电气与电子工程学院教授 e-mail: linxiaomei@ccut.edu.cn

* 通讯作者 e-mail: 1004201519@qq.com; 1124270941@qq.com

耗时较长,可能存在一定的盲目性,并且其精度和收敛速度会因为计算的问题维度过高而受到影响^[13]。本工作选取粒子群算法(PSO)来优化模型参数,避免了一些传统的优化算法在选择参数时存在的问题,获取最佳的惩罚系数 γ 和核函数参数 g 组合,提高了 LIBS 的重复性和检测精度。

1 实验部分

1.1 装置

实验室所搭建的 LIBS 系统如图 1 所示。光源为 Nd:YAG 激光器(Nimma-400 型),波长为 1 064 nm,脉冲宽度为 8 ns,激光能量为 90 mJ。激光器产生的激光经过反射、聚焦透镜聚焦到样品表面,形成激光等离子体。等离子体的光谱信号经光纤探头耦合至光谱仪(Avaspec-2048-USB2)进行分光等处理,实现样品的成分分析。通过控制三维移动平台来移动样品,激光器与光谱仪由数字延时发生器(BNC575,美国 Berkeley Nucleonics Corp)同步控制。

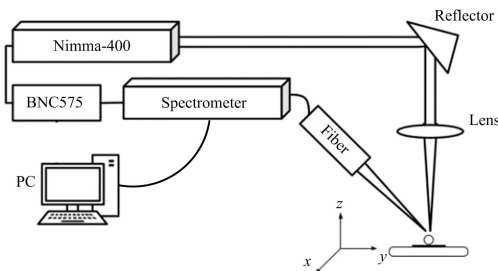


图 1 实验系统原理图

Fig. 1 Schematic diagram of the experimental system

1.2 样品制备

为了确保土壤样品中 Cr 和 Pb 元素的含量稳定,土壤样品是由 $\text{Cr}(\text{NO}_3)_3$ 和 $\text{Pb}(\text{NO}_3)_2$ 的固体结晶和标准土壤样品 GBW07403 混合而成的,加水制成溶液,使分析元素均匀分布于样品中。土样的制作必须要确保样品的干燥,所以在制成溶液过后,需要对溶液进行烘干处理。烘干处理后称取样品 5 g,并将其放进模具,施加的压强为 30 MPa,施压时间为 30 min,最终制成的圆饼直径为 30 mm,厚度为 3 mm。标准土壤样品本身所含分析元素与添加的分析元素相比含量极低,可忽略不计。样品浓度如表 1 所示。

1.3 LSSVM 算法

最小二乘支持向量机(LSSVM)是在支持向量机(SVM)的基础上,由 Suykens 等提出的机器学习方法^[14]。此模型为一个等式约束优化问题,表示为

$$\min J(\omega, e) = \frac{1}{2} \omega^T \omega + \frac{\gamma}{2} \sum_{k=1}^N e_k^2 \quad (1)$$

式(1)中: $\omega = [\omega_1, \dots, \omega_n]^T$ 为权值系数向量; γ 为惩罚系数; e_k 为误差向量; $e_k, 1, 2, \dots, N$ 。根据 Mercer 条件来定义核函数

$$K(x_k, x_l) = \varphi^T(x_k) \varphi(x_l) \quad (2)$$

式(2)中: $\varphi(\cdot) = [\varphi_1(\cdot), \dots, \varphi_n(\cdot)]^T$ 为非线性映射函数; $l=1, 2, \dots, N$ 。最终得到 LSSVM 的非线性模型

$$y(x) = \sum_{k=1}^N \alpha_k K(x, x_k) + b \quad (3)$$

式(3)中: $\alpha_k (k=1, 2, \dots, n)$ 为拉格朗日乘子, $\alpha_k \in R$; b 为偏差向量。

表 1 样品中 Pb 和 Cr 元素的浓度

Table 1 Certified concentration (Wt%) of Pb and Cr in the samples

Sample number	Concentration of Pb and Cr/Wt%
1	0.02
2	0.05
3	0.1
4	0.2
5	0.3
6	0.4
7	0.5
8	0.6
9	0.7
10	0.8
11	0.9
12	1.0

1.4 PSO 算法

PSO 算法将优化问题的所有潜在解视为多维空间中的点,粒子即为不同的点,因此群体由多个粒子构成。粒子会在 PSO 初始化后随机生成,这些随机粒子(随机解)会为了寻找最优解而进行迭代。所有粒子都具有一个自己的矢量速度,最优解就在这些不同粒子之间的相互竞争中被搜寻出来^[15]。粒子迭代优化自身的位置和速度的公式为

$$X_{ik}(t+1) = X_{ik}(t) + V_{ik}(t+1) \quad (4)$$

$$V_{ik}(t+1) = \omega V_{ik}(t) + c_1 r_1 (P_{ik}(t) - X_{ik}(t)) + c_2 r_2 (P_{gk}(t) - X_{ik}(t)) \quad (5)$$

式(5)中: r_1 和 r_2 为随机参数,取值(0, 1)之间; t 为迭代次数; c_1 和 c_2 为学习因子,其数值是非负的; ω 为惯性权重系数; $P_i(t)$ 为个体极值 pBest; $P_g(t)$ 为全局极值 gBest; $k=1, 2, \dots, d$ 。

选取模型预测结果均方根误差 RMSE 作为判别依据,如果预测精度越高,均方根误差的数值就越低。RMSE 的表达式如式(6)

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (X_i - Y_i)^2}{n}} \quad (6)$$

式(6)中: X_i 为实际值, Y_i 为预测值, n 为训练样本个数。

2 结果与讨论

2.1 特征谱线选取

实验获得的样品光谱范围为 205~1 031 nm,若将全光谱数据进行分析,则会导致处理时间过长和分析结果不佳等问题。选择 Pb I 368.35 nm 和 Pb I 405.78 nm 两条相对较好的特征谱线进行对比分析。土壤中含有丰富的元素, Fe

作为基体元素对特征谱线的影响尤为严重。其中 Pb I 368.35 nm 谱线受 Fe I 368.60 nm 影响较大，而 Pb I 405.78 nm 受其他谱线的干扰较小，谱线强度也更高，背景噪声相对较低，故选择 Pb I 405.78 nm 作为 Pb 的分析线。同理，选择 Cr I 425.44 nm，Cr I 427.48 nm 和 Cr I 428.97 nm 三条 Cr 元素特征谱线进行对比分析。Cr 元素的三条谱线与 Fe 的元素特征谱线距离相对较近，其中 Cr I 427.48 nm 附近的 Fe 元素特征谱线最强，受干扰程度最大，而 Cr I 428.97 nm 谱线与其他元素谱线重叠，受干扰严重，所以选择 Cr I 425.44 nm 作为 Cr 的分析线。Pb 和 Cr 元素特征谱线分布如图 2 所示。

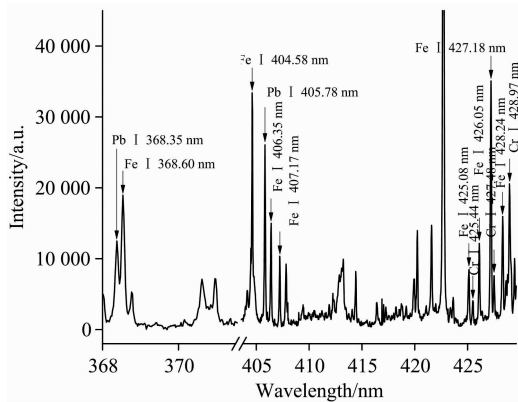


图 2 Pb 和 Cr 元素特征谱线分布

Fig. 2 Characteristic line distributions of Pb and Cr

2.2 定量分析

2.2.1 基于 LSSVM 的定标模型

用 PSO 算法对 LSSVM 优化之前，先测试 LSSVM 算法对定标模型的校准效果。每个浓度的样品采集 20 组数据，将其中的 17 组数据设为训练集，3 组数据设为预测集，构建定标模型。为对该模型的精确度进行评估，故将所有训练集和预测集的数据拟合，拟合效果由相关系数 R^2 来评价。得到的 Pb 拟合曲线如图 3 所示，Cr 拟合曲线如图 4 所示。从拟合曲线中可以看出，Pb 元素 RMSEC 为 0.026 0 Wt%，RMSEP 为 0.101 8 Wt%；Cr 元素 RMSEC 为 0.027 2 Wt%，RMSEP 为 0.078 8 Wt%。Pb 元素定标模型拟合系数 R^2 为

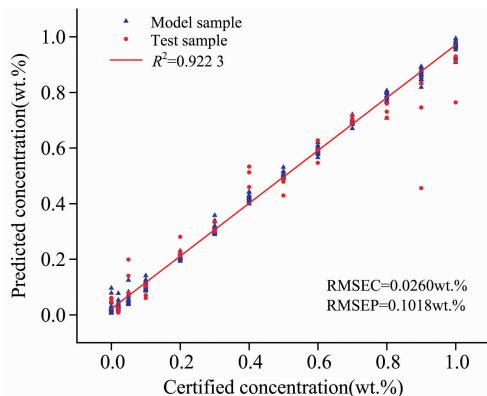


图 3 基于 LSSVM 的 Pb 元素定标模型

Fig. 3 Pb element calibration model based on LSSVM

0.922 3，Cr 元素的 R^2 为 0.948 0，从结果可以看出拟合效果不佳。从图像中可以看出，随着浓度的增加，预测值逐渐低于实际值，这是自吸收效应导致的结果，说明 LSSVM 模型对自吸收效应校准的能力不高，仍需要引入其他算法对该模型进行优化。

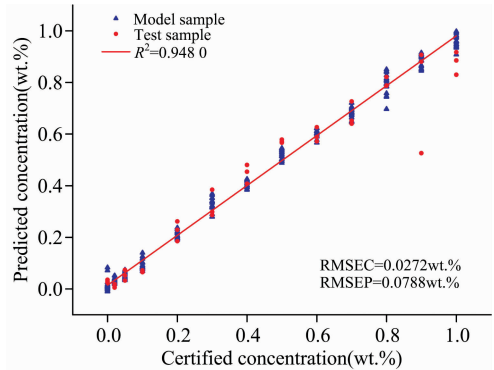


图 4 基于 LSSVM 的 Cr 元素定标模型

Fig. 4 Cr element Calibration model based on LSSVM

2.2.2 基于 PSO-LSSVM 的定标模型

在 LSSVM 模型中，需要先确定参数然后选取核函数，其中需要确定的参数为惩罚系数 γ 和核函数参数 g 。核函数参数 g 主要与泛化能力相关， g 越大，得到的训练模型会越来越平滑，泛化能力越强。 γ 主要与拟合能力相关，惩罚系数 γ 越大，拟合能力越强，但模型的平滑性和泛化能力就会降低。因此需要确定最佳的惩罚系数 γ 和核函数参数 g 组合，使模型的分析效果最佳，故引入 PSO 算法对 LSSVM 模型进行优化。

初始化 PSO 算法的各种参数：学习因子 $c_1 = 1.5$ ， $c_2 = 1.7$ ，迭代次数 $t = 50$ ，种群数量 $sizepop = 20$ 。用 PSO 对训练集进行参数识别，最终得到最佳的惩罚系数 γ 和核函数参数 g 组合，其中 Pb 元素模型的最佳参数组合为 $\gamma_1 = 8\ 096.8$ ， $g_1 = 138.865\ 7$ ；Cr 元素模型的最佳参数组合为 $\gamma_2 = 4\ 908.6$ ， $g_2 = 393.563\ 5$ 。确定最佳参数后，用最佳的参数组合构建 LSSVM 的定标模型。得到的 Pb 拟合曲线如图 5 所示，Cr 拟合曲线如图 6 所示。从图中可以看出，Pb 元素 RMSEC 为 0.022 4 Wt%，RMSEP 为 0.045 8 Wt%；Cr 元

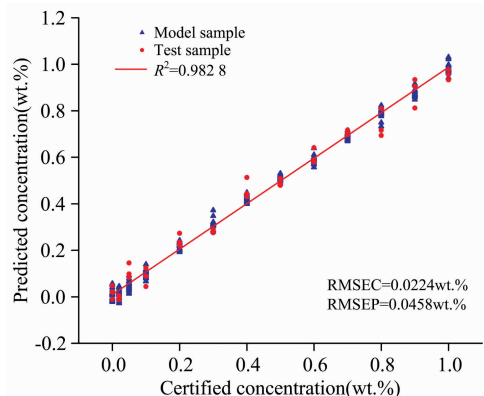


图 5 基于 PSO-LSSVM 的 Pb 元素定标模型

Fig. 5 Pb element calibration model based on PSO-LSSVM

素 RMSEC 为 0.019 1 Wt%，RMSEP 为 0.042 0 Wt%。Pb 元素定标模型拟合系数 R^2 为 0.982 8，Cr 元素的 R^2 为 0.985 0。拟合系数 R^2 的提高十分明显，PSO 算法的优化使 LSSVM 的拟合效果有了很大的提升。同时，RMSEC 和 RMSEP 的数值都有所降低，模型的稳定性和精确度都得到了一定程度的提高，提高了分析精度。两种方法的分析结果如表 2 所示。

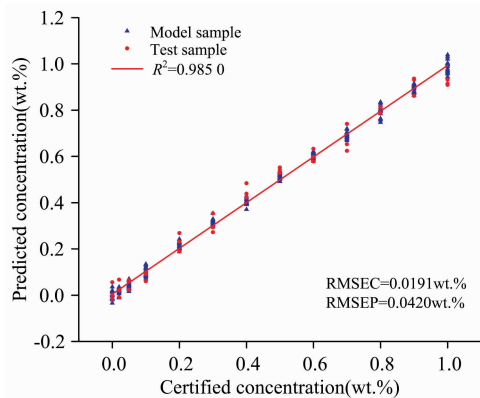


图 6 基于 PSO-LSSVM 的 Cr 元素定标模型

Fig. 6 Cr element calibration model based on PSO-LSSVM

表 2 LSSVM 和 PSO-LSSVM 数据对比

Table 2 LSSVM and PSO-LSSVM data comparison

元素种类	方法	R^2	RMSEC /Wt%	RMSEP /Wt%
Pb	LSSVM	0.922 3	0.026 0	0.101 8
	POS-LSSVM	0.982 8	0.022 4	0.045 8
Cr	LSSVM	0.948 0	0.027 2	0.078 8
	POS-LSSVM	0.985 0	0.019 1	0.042 0

3 结 论

利用 LSSVM 和 PSO-LSSVM 两种方法对土壤中的 Pb 和 Cr 元素建立定标模型。从分析结果可以看出，Pb 和 Cr 元素的 LSSVM 定标模型拟合系数 R^2 只有 0.922 3 和 0.948 0，拟合效果不佳，RMSEC 为 0.026 0 Wt% 和 0.027 2 Wt%，RMSEP 为 0.101 8 Wt% 和 0.078 8 Wt%，说明定标曲线的精确度可以进一步提高。应用 PSO 对 LSSVM 进行优化后，获得了 Pb 和 Cr 最佳惩罚系数 γ 和核函数参数 g 组合，分别为 (8 096.8, 138.865 7) 和 (4 908.6, 393.563 5)。用 PSO-LSSVM 建立 Pb 和 Cr 元素的定标曲线， R^2 分别提高到 0.982 8 和 0.985 0，拟合效果明显提升，RMSEC 降低到 0.022 4 Wt% 和 0.019 1 Wt%，RMSEP 降低到 0.045 8 Wt% 和 0.042 0 Wt%，精确度得到了一定程度的提高。结果表明 LIBS 技术结合 PSO-LSSVM 方法提高检测精度是可行的。

References

- [1] Ilhardt P D, Nuñez Jamie R, Denis E H, et al. Soil Biology and Biochemistry, 2019, 131: 119.
- [2] Kim G, Yoon Y J, Kim H A, et al. Spectrochimica Acta Part B: Atomic Spectroscopy, 2017, 134: 17.
- [3] Villas-Boas P R, Romano R A, Marco Aurélio de Menezes Franco, et al. Geoderma, 2016, 263: 195.
- [4] WANG Ping, LI Dui-yuan (汪萍, 李队员). Spectroscopy and Spectral Analysis (光谱学与光谱分析), 2016, 36(4): 1197.
- [5] De Moraes C P, Barros A I, Júnior, Dário Santos, et al. Microchemical Journal, 2017, 134: 370.
- [6] MENG De-shuo, ZHAO Nan-jing, MA Ming-jun, et al (孟德硕, 赵南京, 马明俊, 等). Journal of Optoelectronics • Laser (光电子 • 激光), 2015, 244(10): 1984.
- [7] ZOU Xiao-heng, HAO Zhong-qi, YI Rong-xing, et al (邹孝恒, 郝中骥, 易荣兴, 等). Chinese Journal of Analytical Chemistry (分析化学), 2017, 2(43): 181.
- [8] Lv Chengxu, Wang Bo, Jiang Xunpeng, et al. Plasma Science and Technology, 2019, 21(3): 112.
- [9] Yang L, Yang S, Li S, et al. Knowledge-Based Systems, 2015, 79(May): 80.
- [10] Zendejboudi, Alireza. Energy Conversion & Management, 2016, 127: 245.
- [11] Torkaman, Mohammad, Safari, et al. Journal of Natural Gas Science & Engineering, 2015, 24: 228.
- [12] Li X, Wu J. Advances in Engineering Research, 2016, 4(1): 1456.
- [13] XU Jun, LU Hai-yan, SHI Gui-juan (许君, 鲁海燕, 石桂娟). Journal of Computer Applications (计算机应用), 2015, 35(3): 668.
- [14] Robledo M A, Sobral H, Garcia V A. Spectrochimica Acta Part B: Atomic Spectroscopy, 2018, 144(6): 7.
- [15] Adrian A M, et al. Ksce Journal of Civil Engineering, 2015, 19(3): 1.

PSO-LSSVM Improves the Accuracy of LIBS Quantitative Analysis

LIN Xiao-mei¹, WANG Xiao-meng¹, HUANG Yu-tao^{1*}, LIN Jing-jun^{2*}

1. Department of Electronics and Electrical Engineering, Changchun University of Technology, Changchun 130012, China

2. Department of Mechanical and Electrical Engineering, Changchun University of Technology, Changchun 130012, China

Abstract Aiming at the problem that the quantitative analysis of soil is greatly affected by the matrix effect and the accuracy of the quantitative analysis of LIBS is not good. The particle swarm algorithm is used to optimize the LSSVM to improve the accuracy of the model. Pb I 405.78 nm and Cr I 425.44 nm was selected as the analysis lines for analysis. Collect the characteristic spectra of twelve samples with different concentrations. The LSSVM calibration model has a low degree of fitting and cannot meet the experimental requirements. The performance of the model needs to be improved. Use particle swarm optimization to optimize the model parameter penalty coefficient γ and kernel function parameter g of LSSVM to obtain the best combination of γ and g . The Pb element is (8 096.8, 138.865 7), and the Cr element is (4 908.6, 393.563 5). Compared with LSSVM, the accuracy of the PSO-LSSVM calibration model is higher. The R^2 of Pb and Cr elements is increased to 0.982 8 and 0.985 0, and the fitting effect is significantly improved. The root means square error of the training set of Pb and Cr elements decreased from 0.026 0 Wt% and 0.027 2 Wt% to 0.022 4 Wt% and 0.019 1 Wt%, and the root means square error of the prediction set was reduced from 0.101 8 Wt% and 0.078 8 Wt% to 0.045 8 Wt% and 0.042 0 Wt%, the stability of the model is further improved. It shows that the PSO-LSSVM algorithm can better reduce the influence of the soil matrix effect and self-absorption effect, and improve the accuracy and stability of the analysis results.

Keywords Laser-induced breakdown spectroscopy; Particle swarm optimization; Least squares support vector machine; Quantitative analysis

(Received Oct. 29, 2020; accepted Feb. 11, 2021)

* Corresponding authors