

DOI: 10.3969/j.issn.1007-5461.2023.03.009

相关向量机结合主成分分析应用于 LIBS技术定量分析

张冉冉, 应璐娜, 周卫东*

(浙江师范大学, 浙江省光信息检测与显示技术研究重点实验室, 浙江 金华 321004)

摘要: 采用相关向量机 (RVM) 结合主成分分析 (PCA) 建立了激光诱导击穿光谱 (LIBS) 技术检测土壤中 Cr 元素含量的定量分析模型。配制了 14 个不同 Cr 元素浓度的土壤样品, 选取其中 10 个作为训练样品集用于构建模型, 另外 4 个作为测试样品集用于模型性能评估。结果表明, 对于土壤中 Cr 元素含量的测量, PCA-RVM 模型的预测精度明显优于 RVM 模型, 整体预测均方根误差由 RVM 模型的 8.00% 减小到 PCA-RVM 模型的 3.21%, 预测精度提高了 59.9%。对测试样品集中全部 4 个待测样品, PCA-RVM 模型多次重复预测结果的相对标准偏差相较于 RVM 模型都显著减小, 且均小于 1.89%, 表明其预测结果具有更好的稳定性。

关键词: 光谱学; 激光诱导击穿光谱; 主成分分析; 相关向量机; 土壤

中图分类号: O657.33

文献标识码: A

文章编号: 1007-5461(2023)03-00376-07

Application of relevance vector machine combined with principal component analysis in quantitative analysis of LIBS

ZHANG Ranran, YING Luna, ZHOU Weidong*

(Key Laboratory of Researching Optical Information Detecting and Display Technology in Zhejiang Province,
Zhejiang Normal University, Jinhua 321004, China)

Abstract: A quantitative analysis model for detecting Cr in soil with laser induced breakdown spectroscopy (LIBS) was established by using correlation vector machine (RVM) combined with principal component analysis (PCA). Fourteen soil samples with different Cr concentrations were prepared, of which ten were selected as training samples for model construction, and the other four as test samples for model performance evaluation. The results show that the prediction accuracy of PCA-RVM model is significantly better than that of RVM model for the measurement of Cr content in soil. The root mean square error (RMSE) of the whole prediction is reduced from 8.00% of RVM model to 3.21% of PCA-RVM model, and the prediction accuracy is improved by 59.9%. Compared with RVM model, the relative standard deviation of repeated prediction results of PCA-RVM model for all four samples in the test

基金项目: 国家自然科学基金 (975186)

作者简介: 张冉冉 (1994 -), 女, 山东人, 研究生, 主要从事激光诱导击穿光谱方面的研究。E-mail: ranranzhangcool@163.com

导师简介: 周卫东 (1966 -), 安徽安庆人, 博士, 研究员, 主要从事激光光谱、激光微纳加工方面的研究。E-mail: wdzhou@zjnu.cn

收稿日期: 2021-04-29; 修改日期: 2021-05-12

*通信作者。

sample set is significantly reduced and is less than 1.89%, indicating that the prediction results of PCA-RVM model have better stability.

Key words: spectroscopy; laser-induced breakdown spectroscopy; principal component analysis; relevance vector machine; soil

0 引言

激光诱导击穿光谱 (LIBS) 技术是一种快速检测样品中元素种类和含量的分析技术, 其不需要对样品进行预处理或仅需进行简单的预处理, 可实现多元素同时探测分析^[1]。LIBS技术的主要机理是利用高能脉冲激光烧蚀样品产生高温等离子体辐射, 获得等离子体辐射中原子和离子的特征谱线, 通过分析特征谱线的位置及强度得出样品中元素的成分及浓度^[2,3]。该技术已广泛应用于食品安全、环境检测、考古鉴定、工业应用等不同领域^[4-8]。

传统的LIBS定量分析方法主要有外标法和内标法^[9,10], 后来研究人员将化学计量学方法与LIBS相结合, 提出了基于偏最小二乘法^[11-13]、人工神经网络^[14]、随机森林^[15,16]、支持向量机^[17-19]、最小二乘支持向量机^[20,21]、相关向量机 (RVM) ^[22] 等模型的LIBS定量分析方法, 这些方法在一定程度上提高了LIBS定量检测性能。其中RVM在2001年由Tipping^[23]首次提出, 它是一种基于有限样本的稀疏概率模型, 该模型不仅拥有支持向量机 (SVM) 的相关优势, 同时在拟合和泛化能力上比SVM更出色, 对噪声的鲁棒性更好^[24,25]。2015年, Yang等^[22]首次把RVM应用于合金钢的LIBS分析, 基于RVM建立的LIBS定量分析模型对合金钢的预测均方根误差为0.71, 预测结果优于分别采用偏最小二乘 (PLS)、人工神经网络 (ANN) 和SVM等方法建立的定量分析模型。Shi等^[26]在对岩石样品的定量分析中, 将主成分分析 (PCA) 结合SVM得到的定量分析模型与仅采用SVM得到的定量分析模型的预测结果相比较, 发现加入PCA方法可以使预测精度提高为原来的7.19倍。Yang等^[27]采用LIBS技术对岩石样品进行分类时, 发现将PCA与SVM相结合不仅能降低噪声、获得较好的精度, 还能有效提高模型的运行效率。

鉴于SVM或RVM与PCA结合在岩石和合金钢的LIBS定量分析应用中均比传统LIBS分析方法有优势^[22-26], 本文将RVM与PCA结合, 应用于土壤中Cr元素的LIBS定量分析检测中, 以提升构建的定量分析模型预测结果的稳定性并提高其预测精度, 并对比分析讨论了主成分分析结合相关向量机 (PCA-RVM) 模型的预测结果与仅使用RVM模型的预测结果。

1 实验部分

1.1 实验装置

实验使用的LIBS装置示意图如图1所示。北京中科思远生产的调Q脉冲Nd:YAG激光器 (型号Plite200-II) 用于激发等离子体, 激光波长为1064 nm, 脉冲宽度为10 ns, 单脉冲能量为100 mJ, 工作频率为1 Hz, 激光器可以同步输出调Q脉冲触发信号, 用于触发光谱仪。脉冲激光光束经过反射镜和焦距为50 mm的凸透镜后聚焦于样品表面下2 mm处。激发的等离子体辐射光谱通过光纤探头耦合导入Avantes公司生产的光纤光栅光谱仪 (型号AvaSpec-2048-USB2), 采集光谱后存储到计算机。光谱仪采用外触发采样工作模式, 响应波长范围为195~510 nm, 采样间隔小于0.08 nm, 积分时间为2 ms。样品置于电动二维平移台上, 采

样时控制移动平台,可使激光在每个烧蚀位置激发采样5次,然后再自动移至样品表面另一个不同的烧蚀位置采样。



图1 实验装置原理框图

Fig. 1 Schematic diagram of experimental setup

1.2 样品制备

4种样品均选自国家标准物质土壤样品(采购于中国计量科学院),编号为GBW07406、GBW07419、GBW07421、GBW07429,并用这些样品按一定比例配制成14个不同的新样品。选取其中10个样品作为训练样品(Train1~Train10),其余4个作为测试样品(#1~#4),每个样品用15 Mpa的压力制成直径约33 mm、厚约2.5 mm的圆片状供实验使用。选取Cr元素作为分析对象,各样品中Cr的浓度以 $\mu\text{g/g}$ 的形式给出,如表1所示。

表1 土壤样品中Cr元素的含量

Table 1 The content of chromium element in soil samples

Content of Cr in train samples/ $(\mu\text{g}\cdot\text{g}^{-1})$										Content of Cr in test samples/ $(\mu\text{g}\cdot\text{g}^{-1})$			
1	2	3	4	5	6	7	8	9	10	#1	#2	#3	#4
75.00	68.00	68.50	81.00	61.00	61.50	74.00	62.00	74.50	87.00	66.00	74.67	70.00	74.33

2 结果与讨论

2.1 数据提取

在实验数据采集过程中,为减小样品表面的不均匀性、激光能量的起伏等因素引起的光谱强度不稳定的影响,对每个样品采样时都控制电动平移台的移动,使激光在每个烧蚀位置激发5次采样,然后再移至样品表面上另一个不同的烧蚀位置采样。采集的每个光谱数据为激光对每个位置激发5次后得到的平均光谱,每个样品采集21个位置,得到21幅光谱数据图,总共收集294幅光谱数据图。

在数据预处理部分,为了降低光谱强度波动性的影响,首先对采集的光谱数据中的谱线强度进行全谱归一化处理,提高光谱数据的信噪比和稳定性^[22]。一幅光谱图中有3994个信息点,一个样品获取21幅光谱图,共得到 3994×21 个光谱信息数据点。为了减少数据处理量,建立预测模型时每个样品选取270.02~278.97 nm的光谱波段,共 105×21 个光谱数据作为输入,以土壤样品Train1 (GWB7406)选取的光谱波段为例,如图2所示。

2.2 主成分分析

原始光谱数据中包含的信息量比较大, 每个光谱信息在不同程度上都与所研究目标元素的浓度有一定的关系, 并且每个光谱信息之间有一定的相关性。换言之, 用来研究目标元素浓度的光谱信息之间有一定程度上的重叠, 重叠部分就称为信息冗余。作为一种化学计量学统计分析方法, PCA^[26]主要是将样本数据映射到一个新的空间, 得到一个新的矩阵代替原始样本数据, 同时消除原始数据间的相关性, 并去除冗余信息。

由图2可见, 所选光谱波段中含有Cr II 274.01 nm、Cr II 274.90 nm、Cr II 275.05 nm、Cr II 275.77 nm四条铬元素的特征谱线, 这四条谱线的强度相对较大, 所以选取单个谱线作为分析线时这几条谱线存在相互干扰的现象, 对Cr元素定量分析的精度会产生较大影响^[27]。训练集的10个样品有 10×21 组光谱数据, 每组105个光谱信息点, 进行主成分分析, 提取出可以代替原数据的主要成分, 从而达到降维目的, 提升模型运行的速率。本研究提取的主成分数目为4, 所使用的4个主成分的累计贡献率达到了95%以上。

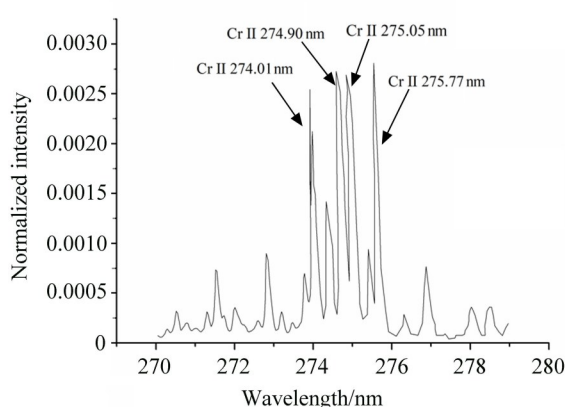


图2 土壤样品Train1 (GWB7406) 在波长270.02~278.97 nm范围内的激光诱导击穿光谱

Fig. 2 LIBS of soil sample Train1 (GWB7406) in the range of 270.02~278.97 nm

2.3 相关向量机

RVM是基于稀疏贝叶斯学习理论的一种从SVM发展而来的统计回归模型^[23]。将PCA与RVM结合建立模型并应用到LIBS技术定量分析中, 可以提高其检测精度和模型的鲁棒性。将主成分分析得到的由4个主成分构成的新的数据信息矩阵作为相关向量机的输入, Cr元素浓度作为相关向量机模型的输出。给定的输入目标矩阵为 $\{I_n, C_n\}$, 其中 I_n 代表主成分分析得到的数据, C_n 代表相应Cr元素的浓度。两者间的关系可以描述为

$$C_n = y(I_n, w) + \varepsilon_n, \quad (1)$$

式中: w 为权重, ε_n 为误差矢量, $y(I_n, w) = \sum_{i=1}^n w_i K(I, I_i) + w_0$, $K(I, I_i)$ 为核函数。核函数是降低计算复杂性的强大工具, 主要目的是将低维空间中的非线性问题转化高维空间中的线性问题。由于数据的非线性关系, 此处选取高斯核函数。对于(1)式, 假设噪声 ε_n 满足0均值和方差为 σ^2 的高斯分布, C_n 也满足高斯分布, 则 $P(C_n | w, \sigma^2) = N(y(I_n, w_n), \sigma^2)$ 。为了避免出现过拟合情况, 对权值 w 的求解上加入先决条件, 即 $P(w_n | \alpha_n) = N(w_n | 0, \alpha)$ 。通过迭代自动求出 w 和 σ^2 后模型构建完成, 模型的定标效果如图3所示, 相关系数为0.99, 可见模型的定标效果比较好。

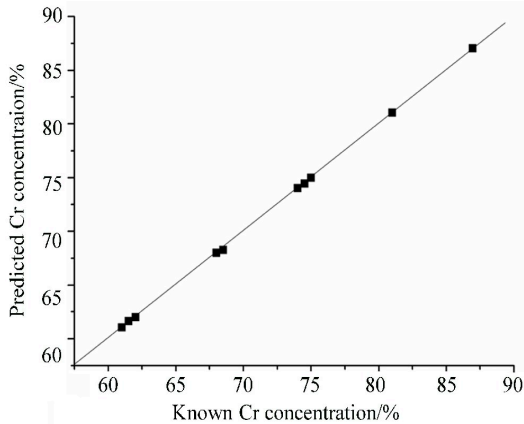


图3 Cr元素定量分析的训练结果

Fig. 3 Training results of quantitative analysis of chromium elements

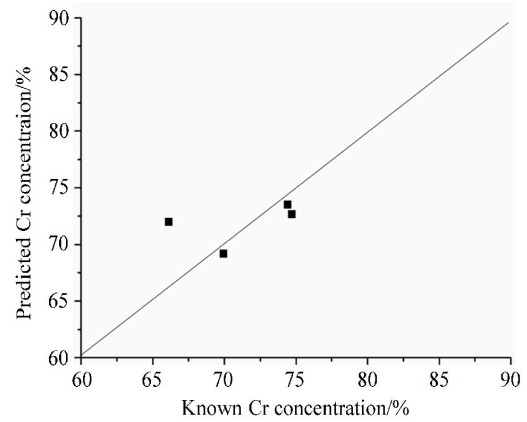


图4 Cr元素定量分析的测试结果

Fig. 4 Testing results of quantitative analysis of chromium elements

2.4 结果对比

将所选取的4个测试样品通过PCA-RVM模型进行预测,测试样本中Cr元素浓度的预测结果如图4所示。此外,选取分析线Cr II 275.05 nm,构建了基于RVM的LIBS定量分析模型,并将PCA-RVM模型和RVM模型的预测结果进行对比分析,如表2、3所示。表2中数据是分别采用两种预测模型对每个测试样品进行21次预测所得到21次预测结果的相对误差平均值(ARE),对测试样本集中全部4个测试样品,PCA-RVM模型预测结果的ARE值都比RVM模型的小。对全部预测结果,PCA-RVM模型的整体预测均方根误差(RMSE)为3.21%,而RVM模型的整体预测均方根误差为8.0%,由此可知PCA-RVM模型相较于RVM模型的预测精度提高了59.9%,表明将主成分分析结合相关向量机构建的PCA-RVM模型的预测精度相较于RVM模型有明显提升。表3列出了两种预测模型分别对每个测试样品进行21次预测所得结果的相对标准偏差,对每一种样品,PCA-RVM模型多次预测结果的相对标准偏差(RSD)均低于RVM模型,且均小于1.89%,表明PCA-RVM模型多次重复预测结果的稳定性优于RVM模型。因此对于选取多组分作为输入的定量分析,PCA-RVM模型能够在一定程度上改善RVM模型的预测精度,并同时提高模型预测结果的可重复性。另外研究还发现PCA-RVM模型的运行速度比RVM模型也有一定程度上的提升。

表2 PCA-RVM和RVM模型对每个样品进行21次预测的平均相对误差(ARE)

Table 2 Average relative error (ARE) of 21 predictions by PCA-RVM and RVM model for each sample

Sample	ARE/%	
	RVM	PCA-RVM
#1	11.03	0.91
#2	6.18	5.86
#3	6.52	2.17
#4	5.73	0.91

表3 PCA-RVM和RVM模型对每个样品进行21次预测的相对标准误差(RSD)

Table 3 Relative standard deviation (RSD) of 21 predictions by PCA-RVM and RVM model for each sample

Sample	RSD /%	
	RVM	PCA-RVM
#1	48.75	1.89
#2	0.91	0.77
#3	1.00	0.19
#4	1.33	0.84

3 结 论

将PCA-RVM分析模型应用于LIBS对土壤元素含量的检测,成功实现了对土壤中Cr元素含量的定量分析。研究表明,与采用RVM构建的模型相比,PCA-RVM模型对土壤Cr元素含量的预测精度有很大提高,整体预测均方根误差由RVM模型的8.00%减小到PCA-RVM模型的3.21%,预测精度提高了59.9%。对全部4个预测样品,PCA-RVM模型多次重复预测的相对标准偏差都小于相应RVM模型的预测值,且均小于1.89%,表明PCA-RVM模型多次重复预测结果的稳定性相较于RVM模型都有提高。PCA算法与RVM相结合可以有效消除光谱信息中的冗余部分,减少了用于RVM模型运行的光谱数据,有利于提升模型的运行速度。

参考文献:

- [1] Cremers D A, Chinni R C. Laser-induced breakdown spectroscopy—Capabilities and limitations [J]. *Applied Spectroscopy Reviews*, 2009, 44 (6): 457-506.
- [2] Aberkane S M, Abdelhamid M, Mokdad F, et al. Sorting zamak alloys via chemometric analysis of their LIBS spectra[J]. *Analytical Methods*, 2017, 9 (24): 3696-3703.
- [3] Ruan F Q, Qi J, Yan C H, et al. Quantitative detection of harmful elements in alloy steel by LIBS technique and sequential backward selection-random forest (SBS-RF) [J]. *Journal of Analytical Atomic Spectrometry*, 2017, 32 (11): 2194-2199.
- [4] Wang Z, Yuan T B, Hou Z Y, et al. Laser-induced breakdown spectroscopy in China [J]. *Frontiers of Physics*, 2014, 9(4): 419-438.
- [5] Ahmad K, Tawfik W, Farooq W A, et al. Analysis of alumina-based titanium carbide composites by laser-induced breakdown spectroscopy [J]. *Applied Physics A*, 2014, 117 (3): 1315-1322.
- [6] Yu Y L, Zhou W D, Su X J. Detection of Cu in solution with double pulse laser-induced breakdown spectroscopy [J]. *Optics Communications*, 2014, 333: 62-66.
- [7] Yuan T B, Wang Z, Li Z, et al. A partial least squares and wavelet-transform hybrid model to analyze carbon content in coal using laser-induced breakdown spectroscopy [J]. *Analytica Chimica Acta*, 2014, 807: 29-35.
- [8] Han D, Joe Y J, Ryu J S, et al. Application of laser-induced breakdown spectroscopy to Arctic sediments in the Chukchi Sea [J]. *Spectrochimica Acta Part B: Atomic Spectroscopy*, 2018, 146: 84-92.
- [9] Herrera K K, Tognoni E, Gornushkin I B, et al. Comparative study of two standard-free approaches in laser-induced breakdown spectroscopy as applied to the quantitative analysis of aluminum alloy standards under vacuum conditions [J]. *Journal of Analytical Atomic Spectrometry*, 2009, 24 (4): 426-438.

- [10] Zhang B H, Jiang Y C, Zhang X Y, *et al.* Quantitative analysis of Mn in soil samples using LIBS [J]. *Spectroscopy and Spectral Analysis*, 2015, 35 (6): 1715-1718.
- [11] Cama-Moncunill X, Markiewicz-Keszycka M, Cama-Moncunill R, *et al.* Sampling effects on the quantification of sodium content in infant formula using laser-induced breakdown spectroscopy (LIBS) [J]. *International Dairy Journal*, 2018, 85: 49-55.
- [12] Zhu Z H, Li J M, Guo Y M, *et al.* Accuracy improvement of boron by molecular emission with a genetic algorithm and partial least squares regression model in laser-induced breakdown spectroscopy [J]. *Journal of Analytical Atomic Spectrometry*, 2018, 33 (2): 205-209.
- [13] Li X W, Wang Z, Lui S L, *et al.* A partial least squares based spectrum normalization method for uncertainty reduction for laser-induced breakdown spectroscopy measurements [J]. *Spectrochimica Acta Part B: Atomic Spectroscopy*, 2013, 88: 180-185.
- [14] Takahashi T, Thornton B. Quantitative methods for compensation of matrix effects and self-absorption in laser induced breakdown spectroscopy signals of solids [J]. *Spectrochimica Acta Part B: Atomic Spectroscopy*, 2017, 138: 31-42.
- [15] Zhang T L, Liang L, Wang K, *et al.* A novel approach for the quantitative analysis of multiple elements in steel based on laser-induced breakdown spectroscopy (LIBS) and random forest regression (RFR) [J]. *Journal of Analytical Atomic Spectrometry*, 2014, 29 (12): 2323-2329.
- [16] Wu S, Zhang T L, Tang H S, *et al.* Quantitative analysis of nonmetal elements in steel using laser-induced breakdown spectroscopy combined with random forest [J]. *Analytical Methods*, 2015, 7 (6): 2425-2432.
- [17] Yan C H, Qi J, Ma J X, *et al.* Determination of carbon and sulfur content in coal by laser induced breakdown spectroscopy combined with kernel-based extreme learning machine [J]. *Chemometrics and Intelligent Laboratory Systems*, 2017, 167: 226-231.
- [18] Yi C C, Lv Y, Xiao H, *et al.* Laser induced breakdown spectroscopy for quantitative analysis based on low-rank matrix approximations [J]. *Journal of Analytical Atomic Spectrometry*, 2017, 32 (11): 2164-2172.
- [19] Gu Y H, Zhao N J, Ma M J, *et al.* Monitoring the heavy element of Cr in agricultural soils using a mobile laser-induced breakdown spectroscopy system with support vector machine [J]. *Chinese Physics Letters*, 2016, 33 (8): 085201.
- [20] Yang J H, Yi C C, Xu J W, *et al.* A laser induced breakdown spectroscopy quantitative analysis method based on the robust least squares support vector machine regression model [J]. *Journal of Analytical Atomic Spectrometry*, 2015, 30 (7): 1541-1551.
- [21] Lin J J, Lin X M, Guo L B, *et al.* Identification accuracy improvement for steel species using a least squares support vector machine and laser-induced breakdown spectroscopy [J]. *Journal of Analytical Atomic Spectrometry*, 2018, 33 (9): 1545-1551.
- [22] Yang J H, Yi C C, Xu J W, *et al.* Laser-induced breakdown spectroscopy quantitative analysis method via adaptive analytical line selection and relevance vector machine regression model [J]. *Spectrochimica Acta Part B: Atomic Spectroscopy*, 2015, 107: 45-55.
- [23] Tipping M E. Sparse Bayesian learning and the relevance vector machine [J]. *Journal of Machine Learning Research*, 2001, 1 (3): 211-244.
- [24] Wipf D P, Rao B D. Sparse Bayesian learning for basis selection [J]. *IEEE Transactions on Signal Processing*, 2004, 52 (8): 2153-2164.
- [25] Ji S H, Xue Y, Carin L. Bayesian compressive sensing [J]. *IEEE Transactions on Signal Processing*, 2008, 56 (6): 2346-2356.
- [26] Shi Q, Niu G H, Lin Q Y, *et al.* Quantitative analysis of sedimentary rocks using laser-induced breakdown spectroscopy: Comparison of support vector regression and partial least squares regression chemometric methods [J]. *Journal of Analytical Atomic Spectrometry*, 2015, 30 (12): 2384-2393.
- [27] Yang H X, Fu H B, Wang H D, *et al.* Laser-induced breakdown spectroscopy applied to the characterization of rock by support vector machine combined with principal component analysis [J]. *Chinese Physics B*, 2016, 25 (6): 065201.