

基于 K-CV 参数优化支持向量机的 LIBS 燃煤热值定量分析

董美蓉^{1,2,3}, 韦丽萍^{1,2,3}, 陆继东^{1,2,3*}, 黎文兵^{1,2,3}, 陆盛资^{1,2,3},
黄健伟^{1,2,3}, 李诗诗^{1,2,3}, 骆发胜^{1,2,3}, 聂嘉朗¹

1. 华南理工大学电力学院, 广东 广州 510640
2. 广东省能源高效低污染转化工程技术研究中心, 广州 广东 510640
3. 广东省能源高效清洁利用重点实验室, 广东 广州 510640

摘要 热值是煤质特性的重要参数之一,很大程度上影响着燃煤锅炉的运行。为了克服传统检测方法所存在的问题,将激光诱导击穿光谱(LIBS)应用于燃煤热值的定量分析。煤的结构复杂,所含的元素种类众多,包括了主量元素、微量元素和痕量元素,致使煤的 LIBS 光谱信息复杂。如何有效提取 LIBS 光谱信息,实现准确的量化测量是 LIBS 在煤特性检测中发挥作用的前提和基础。近年来,随着人工智能技术的发展,相关的分析技术也开始应用于煤的工业指标分析和热值预测中。为实现煤样品中 LIBS 光谱信息的有效提取,同时为克服常规的分析方法易出现的过渡拟合、收敛性不好等问题,提出采用结合 K-fold Cross Validation(K-CV)参数优化的支持向量机(SVM)回归方法,实现 LIBS 定量分析煤中的热值。SVM 方法是结构风险最小化的近似实现,可用于模式分类和非线性回归。为了得到有效的 LIBS 分析模型,实验选用 44 种电厂常用的热值含量不同的煤样作为实验对象,选择其中 33 个作为训练集,剩余 11 个为测试集。利用搭建的 LIBS 实验系统获取所选煤样品的等离子体发射光谱数据,首先分析了 SVM 热值回归模型的参数-惩罚因子 C 、核函数参数 g 与模型精度的关联,确定 C 和 g 最佳取值范围,然后分别建立了基于 LIBS 全谱和某些元素(非金属元素和金属元素)特征光谱的 SVM 回归模型。利用训练集光谱数据,结合 K-CV 法得到热值 SVM 回归模型的最优参数 C 和 g 的值,建立基于 SVM 最优参数的煤热值定量分析模型。然后将测试集的光谱数据作为输入量用于测试所建立模型的可靠性,得到分别采用全谱、非金属元素特征光谱、非金属与金属元素特征谱相结合的热值定量分析模型,其决定系数 R^2 均达到 0.99 以上,均方误差分别为 0.12, 0.17 和 0.06 ($\text{MJ} \cdot \text{kg}^{-1}$)², 预测平均相对偏差分别为 1.2%, 1.23% 和 0.69%。结果表明:基于 K-CV 参数优化 SVM 回归方法可用于 LIBS 技术实现燃煤热值的定量分析,且可得到较高的分析精确度和准确度;同时通过对比选用不同的光谱特征的定量分析模型可知,采用非金属与金属元素的特征光谱所建立的基于 K-CV 参数优化 SVM 的热值定量模型,能够有效提高 LIBS 应用于快速检测煤热值的精度和准确度,实现煤热值的准确预测。

关键词 煤; 热值; 激光诱导击穿光谱; 支持向量机; 定量; K-CV

中图分类号: O433.1 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2019)07-2202-08

引言

煤的热值是计算煤耗、热效率、燃烧设备热平衡、确定不同煤混和配比等的关键参数,也是评价煤品质的重要指标。目前,煤热值的离线分析方法主要是通过弹筒量热仪测

量反应物与生成物的焓差计算得到^[1]。这种方法操作简单、结果比较精确,但是整个分析周期一般需要几个小时,难以满足工业生产过程的要求。

市场上相对较成熟的可实现包括热值分析的煤工业分析指标快速分析装置主要是基于 γ 射线和中子活化分析(PG-NAA)的煤质分析仪^[2]。但该技术中 γ 射线对人体存在很大

收稿日期: 2018-05-24, 修订日期: 2018-10-13

基金项目: 国家自然科学基金项目(51476061, 51406059), 广东省自然科学基金重点项目(2017B030311009), 中央高校基本科研业务费专项资金(华南理工大学)项目(2017ZD027)资助

作者简介: 董美蓉, 女, 1985 年生, 华南理工大学电力学院副教授 e-mail: epdongmr@scut.edu.cn

* 通讯联系人 e-mail: jdliu@scut.edu.cn

的安全隐患,且该类设备复杂,投资大,不利于市场推广。因此,应用新的理论和方法实现燃煤热值的快速测定具有重要意义。激光诱导击穿光谱(laser-induced breakdown spectroscopy, LIBS)是近年来发展起来的一种可实现样品中多组分同步快速分析技术,已被尝试应用于各种工业过程的质量控制和过程监测^[3]。

姚顺春等^[4]运用 LIBS 分析燃煤灰成分组成元素,结合内标法实现灰分主要成分(SiO_2 , Al_2O_3 , Fe_2O_3 , CaO , MgO 和 TiO_2)的定量分析。张雷等^[5]运用 LIBS 实现了煤中各元素(C, Ca, Mg, Ti, Si, H, Al, Fe)的定量分析,并由元素分析得到了工业分析,该团队^[6]还利用 LIBS 分析了飞灰中的未燃碳含量。Haider 等^[7]运用 LIBS 检测了煤中的 C 和 Si 以及一些大都存在于灰分中的金属元素(Fe, Ti, Al, Ca, Na, Cu, Zr, Nd, Yb, Ce, Sm 和 Dy 等)。董美蓉等^[8]运用 LIBS 结合多元定标法定量分析了煤粉碳元素,得到预测值的相对误差在 5% 以内。

LIBS 作为一项测量技术,基体效应明显,探测到的光谱信息可以一定程度上反应基体特征。因此 LIBS 不仅可实现煤中各元素分析,还可实现煤的工业指标分析。而如何有效提取 LIBS 光谱信息,实现准确的量化测量是 LIBS 技术在煤特性检测中发挥作用的前提和基础,也是目前国内外研究者所关注的热点问题。近年来,随着人工智能技术的发展,相关的分析技术也开始应用于煤的工业指标分析和热值预测中。Haider 等^[9]采用 LIBS 结合线性回归方法实现煤中灰分分析,其相关系数达到 0.99。Yao 等^[10]采用 LIBS 结合偏最小二乘法定量分析煤中灰分,其预测值的平均相对标准偏差为 8.50%。同时还采用 LIBS 结合人工神经网络、遗传算法定量分析煤中的灰分、挥发分、固定碳和热值,取得了不错的结果^[11-12]。黎文兵等^[13]结合 LIBS 光谱平滑预处理和偏最小二乘法,实现煤中热值的定量分析,提升了煤中热值的预测可靠性。Wang 等^[14-15]以煤样品的 LIBS 光谱数据为基础,用偏最小二乘法对煤的灰分、挥发分及热值进行定量分析,提高了模型预测精度。王帝等^[16]提出了 LIBS 对煤中热值检测的一种新型校正模型,对 LIBS 光谱数据进行自吸收效应、元素间相互干扰、基体效应修正后再进行多元线性

回归建模,提高了煤中热值检测的可靠程度、预测能力。

常规的神经网络方法存在局部极小化、过学习、收敛速度慢等缺陷,而偏最小二乘法存在主成分选取非最优化和过渡拟合、收敛性不佳以及收敛速度慢等问题,减弱了其模型的预测能力。谭鹏等^[17]引入支持向量机(SVM)方法应用在基于工业分析基础上的煤热值预测,克服了以上方法中存在的问题,提高了煤热值的预测精度。另外,极少数作者涉及到了支持向量机模型的参数优化问题,如江文豪等^[18]提出了一种基于遗传算法优化参数支持向量机(GA-SVM)的燃煤热值预测方法,提高了模型的逼近能力,在学习能力和泛化能力之间取得了平衡。但 GA 算法存在不成熟的过早收敛或者收敛缓慢甚至不收敛等问题,其搜索效果不稳定。而基于 K-fold Cross Validation(K-CV)参数优化方法可以有效避免上述问题,且 K-CV 还可有效地避免过学习以及欠学习状态的发生,最后得出的结果具有较强的说服力。

本工作采用基于 K-CV 参数优化支持向量机算法对煤的热值进行定量分析。首先分析了核函数的参数和惩罚因子对 SVM 热值预测模型的影响,然后分别建立了基于全谱和特征元素光谱特征(金属元素和非金属元素)的 SVM 定量分析模型,并比较了基于不同光谱特征所建立模型的准确度和精确度。

1 实验部分

1.1 样品

实验选取了 44 种电厂常用煤样品,经过磨煤、过筛(30 个 $\cdot\text{cm}^{-2}$ 孔),最后烘制成干燥基,测到的样品在干燥基下的高位热值如表 1 中所示(数据由广州市特种承压设备检测研究院提供),并将 44 个煤样本分为训练集和测试集两组。为了使测试集的热值等梯度地分布于训练集中,选取 #3, #7, #11, #15, #19, #23, #27, #31, #35, #39 和 #43 这 11 个样品作为测试集样品,剩下的 33 个煤样品为训练集样品。为了提高 LIBS 分析系统信号采集的稳定性,实验前将各煤样称取 3.5 g,使用压片机在压力为 6.5 t 下将粉末状煤样品压成直径 3 mm、表面平整的圆柱状样品。

表 1 44 种煤样品的高位热值(干燥基)

Table 1 Heat value of 44 coal samples (dry basis)

样品	热值/(MJ $\cdot\text{kg}^{-1}$)	样品	热值/(MJ $\cdot\text{kg}^{-1}$)	样品	热值/(MJ $\cdot\text{kg}^{-1}$)	样品	热值/(MJ $\cdot\text{kg}^{-1}$)
#1	18.70	#12	23.36	#23	26.47	#34	29.16
#2	20.48	#13	23.81	#24	26.63	#35	29.49
#3	21.05	#14	23.85	#25	26.96	#36	29.62
#4	21.65	#15	24.29	#26	27.04	#37	30.28
#5	21.77	#16	24.86	#27	27.29	#38	30.58
#6	21.86	#17	25.10	#28	27.30	#39	30.76
#7	22.02	#18	25.13	#29	27.59	#40	30.78
#8	22.03	#19	25.31	#30	27.98	#41	30.86
#9	22.50	#20	25.96	#31	28.01	#42	31.49
#10	22.59	#21	26.40	#32	28.43	#43	31.50
#11	22.59	#22	26.44	#33	28.47	#44	32.58

1.2 仪器及参数

采用的 LIBS 实验系统, 详见参考文献[13], 该系统中的激光光源是波长为 1 064 nm 的 Nd: YAG 调 Q 固体脉冲激光器 (Brilliant EaZy, 法国 Quantel 公司), 激光的最大能量是 300 mJ, 脉宽是 6 ns。实验过程中设定激光器的脉冲频率为 1 Hz、激光能量为 50.75 mJ (标准偏差 0.61%)。激光从激光器输出, 经 45° 反射镜、焦距为 100 mm 的透镜垂直聚焦在煤样品上 (焦点位置在样品表面以下 1 mm)。集光系统是内部集成了 2 048 pixels 线阵 CCD 作为探测器件的光纤光谱仪 (AvaSpec-2048FT, 荷兰 Avantes 公司), 其由两个探测通道组成, 波长覆盖范围为 230~400 和 580~790 nm, 分辨率为 0.3~0.4 nm。激光与样品作用后产生的等离子体由与样品表面成 45° 的准直透镜收集, 经光谱仪、CCD, 实现分光、光电转换、数据的采集和储存。为获得较好的信噪比, 激光器触发与光谱仪开始采集之间的延时通过 DG535 脉冲信号发生器设置为 1.664 μs, CCD 的积分时间设置为 1.1 ms (最小积分时间)。每个煤样品选取 6 个位置, 每个位置采集 50 幅光谱, 取平均后作为一个样品的 LIBS 数据。

1.3 SVR 理论

支持向量机 (SVM) 是结构风险最小化的近似实现, 其能够根据有限的样本信息, 在学习能力和泛化能力之间取得平衡。SVM 可用于模式分类和非线性回归, 当将其用于回归时被称为支持向量机回归 (SVR)。常用的支持向量回归机有 ϵ -SVR 支持向量回归机和 ν -SVR 支持向量回归机, 使用 ϵ -SVR 需要先确定不敏感损失函数中的参数, 而在煤的热值预测中却是很难确定该参数。本文运用到的 SVM 回归机类型是 ν -SVR 支持向量回归机, 该方法可以通过选择参数 ν 自动确定 ϵ 的值, 使得算法更加适用于煤中热值的定量分析。该方法的基本思想是: 获得煤样品的训练集样本 (x_i, z_i) (其中 $i=1, 2, \dots, l$; $x_i \in R^m$, 输入量即煤样品光谱强度矩阵; $z_i \in R$, 为对应的输出值即煤的热值, 利用核函数将 x 映射到一个高维空特征空间, 在该特征空间用以下回归函数进行回归

$$f(X) = \omega \varphi(x) + b \quad (1)$$

其中, ω 和 b 分别是权向量、偏差; $\varphi(x)$ 支持向量机通过对以下目标函数极小化求回归函数

$$\min_{\omega, b, \xi, \xi^*, \epsilon} \frac{1}{2} \omega^T \omega + C \left(\nu \epsilon + \frac{1}{l} \sum_{i=1}^l (\xi_i + \xi_i^*) \right) \quad (2)$$

其中, ν 是控制支持向量机数量的参数, $\nu \in (0, 1]$ 。式 (2) 约束条件为

$$\begin{aligned} (\omega^T \varphi(x_i) + b) - z_i &\leq \epsilon + \xi_i \\ z_i - (\omega^T \varphi(x_i) + b) &\leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* &\geq 0, i = 1, \dots, l, \epsilon \geq 0 \end{aligned} \quad (3)$$

其中, C 为惩罚因子, $C > 0$; $\xi_i, \xi_i^* \geq 0$ 为松弛因子; ϵ 为不敏感损失函数。上述最优化问题的对偶问题是

$$\min_{\alpha, \alpha^*} \frac{1}{2} (\alpha - \alpha^*)^T Q (\alpha - \alpha^*) + z^T (\alpha - \alpha^*) \quad (4)$$

约束条件是

$$\begin{aligned} e^T (\alpha - \alpha^*) &= 0, e^T (\alpha + \alpha^*) \leq C\nu \\ 0 &\leq \alpha_i, \alpha_i^* \leq \frac{C}{l}, i = 1, \dots, l \end{aligned} \quad (5)$$

求解以上对偶问题后可以得到

$$f(x) = \sum_{i=1}^l (-\alpha_i + \alpha_i^*) K(x_i, x) + b \quad (6)$$

式 (6) 中, $K(x_i, x)$ 为核函数。由于煤样品的 LIBS 光谱数据包含信息众多, 光谱数据与煤中热值存在非线性的对应关系; 同时所探测到的光谱数据不仅包含了各主要元素的特征谱线, 还包含了噪声、背景等信息。致使所采用的核函数需适用于非线性回归且对数据中存在的噪声有较好的抗干扰能力, 因此选择高斯径向基核函数 (RBF)

$$K(x, x_i) = \exp(-g \|x - x_i\|^2), g > 0 \quad (7)$$

其中, g 为核函数参数, $g = \frac{1}{2\sigma^2}$, σ 为核函数宽度参数, 控制核函数的径向作用范围。

影响 ν -SVR 模型的参数有惩罚因子 C , 核函数参数 g , 控制支持向量数量的参数 ν 。建立的 ν -SVR 模型采用的核函数是 RBF 核函数, 在利用该模型对煤中热值进行回归预测时, 较大程度上影响该模型预测准确度和精度的参数是 C 和 g 。 C 的大小表示当超出经验误差后对样本的惩罚程度, 该值越大, 则惩罚程度越高, 但是该值过大, 会导致出现过拟合状态。而当 C 值过小, 则说明对超出经验误差的样本惩罚过小, 会导致欠拟合状态。 g 的大小决定核函数的特征空间, 可用于表征样本数据子空间分布的复杂程度, 如果该值过小, 模型会相对比较复杂, 推广能力得不到保证。而 g 过大, 模型又难以达到足够的精度。

1.4 K-CV 参数优化理论

K-CV 参数优化方法是基于交叉验证 (cross validation, CV) 思想下的一种网格寻优方法。K-CV 参数优化方法的原理是将样本分成 K 组, 将每组子集的样本分别作一次测试集, 而同时, 将其余的 $K-1$ 组子集的样本作为训练集, 共可获得 K 个模型, 将 K 个模型对应测试集的回归均方误差的均值作为回归性能判断指标。对 33 个训练集样品的全谱, 结合 K-CV 法对模型参数进行优化, 得到的最低均方误差 (CVmse) 随 K 值的变化如表 2 所示, 从中可以看出在 $K=5$ 时最低均方误差最小, 因此选取 $K=5$ 。

表 2 参数优化过程的最低均方误差
随 K 值的变化结果

Table 2 The lowest mean square error (CVmse) varied with the K value in the parameter optimization process

K	CVmse/(MJ · kg ⁻¹) ²
2	9.13
3	0.39
4	0.41
5	0.21
6	0.23
7	0.26
8	0.22
9	0.23
10	0.25

1.5 模型评价指标

基于 K-CV 参数优化 SVM 回归模型的评价指标包括：首先从模型整体评价其预测能力，参数优化 CV 过程中最低均方误差 (best cross validation MSE, CVmse)、均方误差 (mean square error, MSE)、决定系数 (R^2)；其次从单次测量上评价模型的预测能力，预测平均相对偏差 (average relative deviation, ARD)、预测最大相对偏差 (maximum relative deviation, MRD)，这里的相对偏差指煤样品测试集热值预测值与真值的相对偏差的绝对值，其表达式为

$$\text{相对偏差} = \left| \frac{\text{单次预测值} - \text{真值}}{\text{真值}} \right| \times 100\% \quad (8)$$

均方误差的表达式为

$$\text{MSE} = \frac{1}{k} \sum_{i=1}^k (y_i - \tilde{y}_i)^2 \quad (9)$$

其中, y_i 和 \tilde{y}_i 为第 i 个煤样品热值的真实值、预测值, k 为样本数。

R^2 的表达式为

$$R^2 = \frac{\left(\sum_{i=1}^n (y_i - \bar{y})(\tilde{y}_i - \bar{\tilde{y}}) \right)^2}{\sum_{i=1}^n (y_i - \bar{y})^2 \sum_{i=1}^n (\tilde{y}_i - \bar{\tilde{y}})^2} \quad (10)$$

其中, y_i , \tilde{y}_i , \bar{y} 和 $\bar{\tilde{y}}$ 为第 i 个煤样品热值的真实值、预测值、样品热值真实值的平均值、样品热值预测值的平均值。

2 结果与讨论

在对煤样品热值进行回归建模前，无论是利用全谱还是特征谱(非金属谱、非金属加金属谱)建立 SVM 回归模型，都对光谱数据进行平滑预处理^[13]。图 1 是 #1 煤样品经过平滑处理后的 LIBS 光谱图。

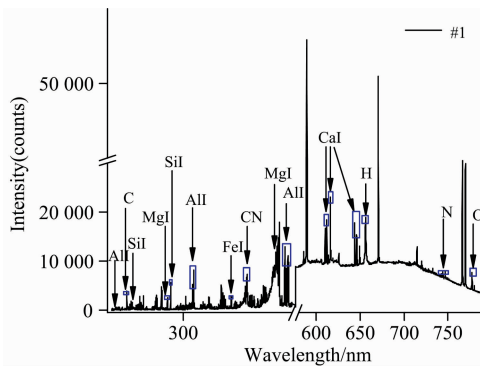


图 1 #1 煤样品平滑处理后的 LIBS 光谱图

Fig. 1 LIBS spectra of 1# coal sample after smoothing

煤的基体成分复杂，在受激发形成等离子体后冷却过程中，不同元素发出的原子光谱频率和强度不同，则每一个样品不同波长点(即不同维度)的光谱强度数值量纲差别较大。为了防止归一化过程中量纲数量级低的属性变为零，从而使样本某些原始信息丢失，对平滑处理后的光谱矩阵每一维度进行归一化(即按列归一化)。具体步骤为：从 m 个热值不同的煤样品中获取 m 组光谱数据；将 m 组光谱数据都运用

式(11)对每一维度归一化处理，得到 $[y_{\min}, y_{\max}]$ ，则归一化处理后的每一维度里的 y 可表示为

$$y = \frac{(y_{\max} - y_{\min})(x - x_{\min})}{(x_{\max} - x_{\min})} + y_{\min} \quad (11)$$

式中, x 为原始数据, y 为归一化后的数据, y_{\max} 为归一化范围的上限值, y_{\min} 为归一化范围的下限值, x_{\max} 为每一维度里最大的原始值, x_{\min} 为每一维度里最小的原始值。

2.1 基于全谱的 SVR 热值预测分析

为了研究支持向量机回归模型参数-惩罚因子 C 、核函数参数 g 对热值回归模型的影响，首先将探测到的全谱数据作为输入变量，取搜索步长为 2^n , $n \in [-12, 12]$ 分析参数 C 和 g 在取不同数值(即 C 的取值是 $2^{-12}, 2^{-11}, \dots, 2^{12}$, g 的取值是 $2^{-12}, 2^{-11}, \dots, 2^{12}$)的情况下，对训练模型的影响。如图 2 中的 (a) 所示为 C 和 g 的变化对训练模型均方误差 (MSE) 的影响，从中可以看出，在 g 值小于 2^{-1} 时, C 值越大训练模型的 MSE 越小。而在 C 值大于 2^3 后, g 值得变化对模型结果的影响不大。

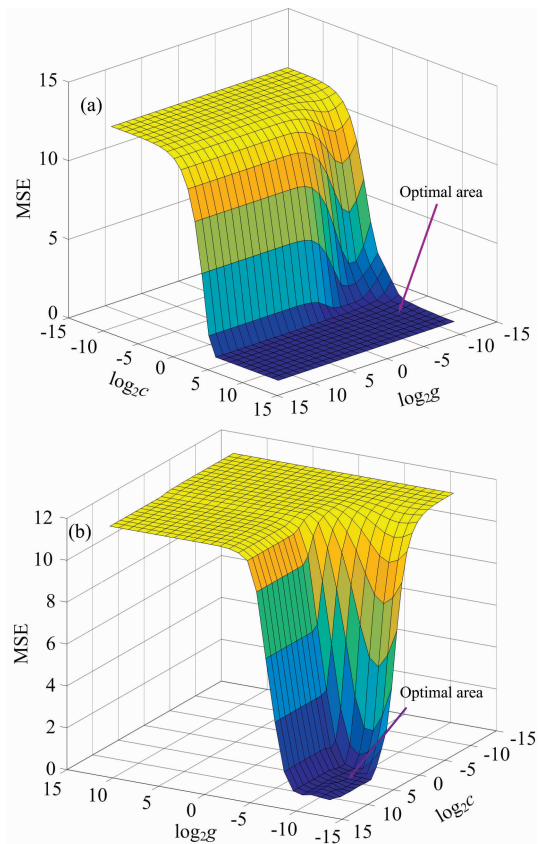


图 2 C 和 g 的变化对训练模型均方误差 (MSE) (a) 与模型预测精度 (b) 的影响

Fig. 2 Effect of changes in C and g on (a) MSE of training model and (b) model prediction accuracy

如图 2 的 (b) 所示为 C 和 g 的变化对测试集均方误差 (MSE) 的影响，旨在进一步分析参数 C 和 g 不同时对模型预测热值精度的影响。从中可以看出，当 C 值超过 1024 (即 2^{10}) 时，即使在 g 值很小的情况下，测试集的 MSE 也会随着

C 值的增大而增大, 即会产生过拟合现象(拟合的模型完美匹配训练集数据, 但对新数据的预测结果较差)。而在 C 值大于 2^5 小于 2^{10} 、g 值小于 2^{-10} 时, 再减小 g 值取值(比如从 2^{-10} 到 2^{-12} 这段), 对模型的影响比较微小。因此, 在对煤热值进行支持向量机回归模型参数优化时, 惩罚因子 C 的寻优范围可选择小于 2^{10} , g 值得寻优范围可考虑大于 2^{-12} 。即在基于全谱或特征谱建立热值定量模型过程中, C 值的优化区间为: $[2^{-10}, 2^{10}]$, g 值的优化区间为 $[2^{-12}, 2^{12}]$ 。

利用实验采集到的全部光谱来建立热值预测模型。运用

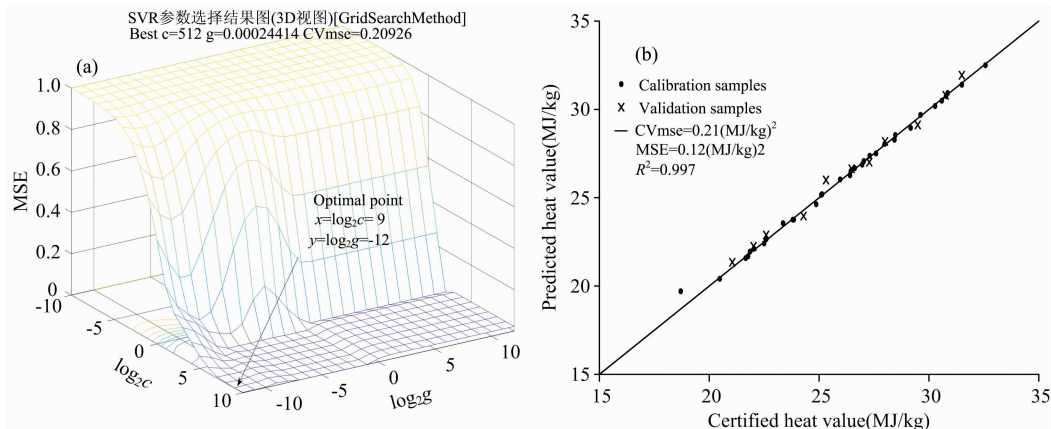


图 3 基于全谱的参数优化结果(a)与热值预测模型(b)

Fig. 3 Full-spectrum-based parameter optimization results (a) and heat value predicted model (b)

表 3 分别利用全谱、特征谱预测热值的结果对比

Table 3 Comparison of predict heat values between full spectrum and characteristic spectrum

真值/ (MJ · kg ⁻¹)	全谱		非金属谱		非金属加金属谱	
	预测值/ (MJ · kg ⁻¹)	相对偏差/ %	预测值/ (MJ · kg ⁻¹)	相对偏差/ %	预测值/ (MJ · kg ⁻¹)	相对偏差/ %
21.05	21.36	1.47	21.46	1.95	21.02	0.14
22.02	22.24	1.00	21.92	0.45	22.00	0.09
22.59	22.89	1.33	23.20	2.70	22.66	0.31
24.29	23.95	1.40	24.24	0.21	24.36	0.29
25.31	26.02	2.81	25.97	2.61	25.94	2.49
26.47	26.65	0.68	26.77	1.13	26.76	1.10
27.29	27.01	1.03	27.53	0.88	27.16	0.48
28.01	28.20	0.68	28.17	0.57	28.17	0.57
29.49	29.10	1.32	29.47	0.07	29.26	0.78
30.76	30.79	0.10	30.89	0.42	30.61	0.49
31.50	31.94	1.40	32.30	2.54	31.76	0.83

2.2 基于特征谱线的 SVR 热值分析

C 和 H 元素对燃煤热值有直接影响, O 和 N 是有机物中的不可燃成分, 它们的含量对煤的热值也有间接影响。因此可考虑利用所探测到的与这些非金属元素相关的特征光谱 C, H, CN, O 和 N 作为 SVM 模型的输入变量, 用以建立热值分析模型, 以进一步分析 LIBS 光谱与预测煤中热值精度的关联性。采用的非金属元素相关的特征谱线如表 4 所示。

运用该方法建立预测模型, 得出的最佳参数 C 为 128, g 为 0.5, 交叉验证过程中 CVmse 为 $0.52 (\text{MJ} \cdot \text{kg}^{-1})^2$ 。运用该最优模型预测测试集煤样品的热值, 预测结果如图 4(a)所

示。训练集的 R^2 为 0.997, 测试集的 R^2 为 0.993, 且 MSE 为 $0.17 (\text{MJ} \cdot \text{kg}^{-1})^2$ 。预测值与真值的对比分析如表 3 所示, 最大相对偏差为 2.7%, 最小相对偏差为 0.07%, 平均相对偏差 ARD 为 1.23%。

除了 C, H, O, N 几种非金属元素对燃煤热值有影响外, 煤中灰分的主要成份 Si, Al, Fe 和 Ca 以及少量 Mg 等对热值也会产生影响。因此考虑将所采集到的各主要非金属元素和金属元素(C, H, CN, O, N, Ca, Mg, Al, Fe, Si)的特征光谱强度同时作为输入变量, 所选择的特征峰的谱线如表 4 所示。

表 4 主要非金属与金属元素的特征谱线
Table 4 Spectral emission lines of the major non-metal and metal elements

元素	波长/nm					
C	247.869 3					
H	656.350 3					
CN	358.555 4	358.172 8				
O	777.335 2					
N	742.415 0	744.325 7	746.890 9			
Ca I	610.290 2	612.205 4	616.260 5	643.956 7	646.247 4	649.441 8
Mg I	285.214 7	383.320 7	383.897 9			
Al I	236.682 2	237.322 3	308.196 6	309.284 0	394.421 3	396.169 5
Fe I	248.321 3	249.044 0	252.379 4	278.787 6	344.039 6	358.172 8
	373.526 7	374.560 9	374.855 9	375.886 9	382.091 2	388.698 6
Si I	250.668 1	251.929 4	252.829 2	288.149 5		

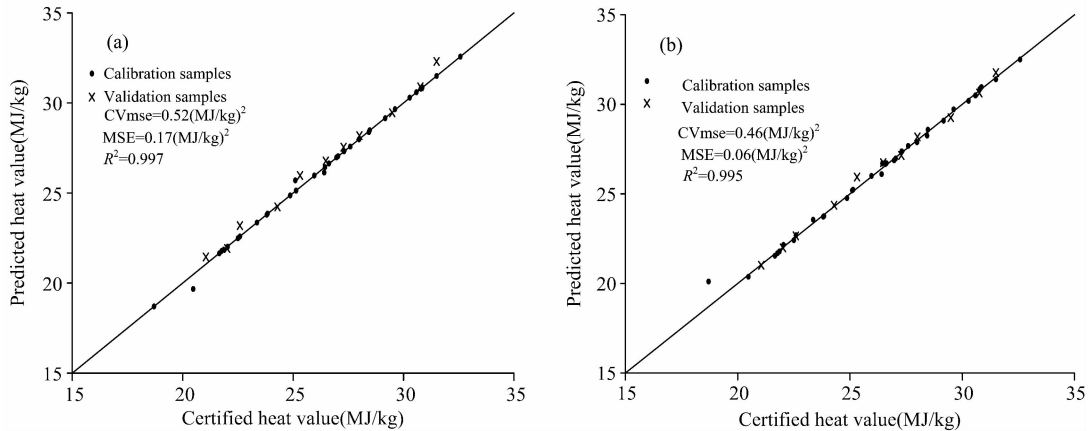


图 4 基于非金属特征光谱 (a) 和非金属与金属结合特征光谱 (b) 的热值模型

Fig. 4 Heat value models based on Based on non-metal spectra (a) and non-metal plus metal spectra (b)

基于所述非金属与金属元素的特征谱线所建立的定量分析模型, 得出的最佳参数 C 为 1.024, g 为 0.015 625, 交叉验证过程中 $CVmse$ 为 $0.46 (MJ \cdot kg^{-1})^2$ 。运用该最优模型预测测试集煤样品的热值, 预测结果如图 4 中 (b) 所示。训练集的 R^2 为 0.995, 测试集的 R^2 为 0.995, 且 MSE 为 $0.06 (MJ \cdot kg^{-1})^2$ 。预测值与真值的对比分析如表 3 所示, 最大相对偏差为 2.49%, 最小相对偏差为 0.09%, 平均相对偏差 ARD 为 0.69%。

图 5 为基于全谱和特征元素光谱所建立的热值定量分析模型对训练集样本的模型评价指标的对比。从中可以看出, 利用非金属元素光谱强度建立的热值 SVM 回归模型可以达到较高的分析精度, 主要是因为热值主要与煤中有机质有关, 但是相对利用其他两种输入变量的定量分析方法而言, 其均方误差 MSE 、平均相对偏差 ARD 较大, 主要是因为非金属元素的激发能级较高, 同时易受环境的影响, 因此谱线的波动会较大。

在燃烧过程中, 煤中的矿物质也会发生分解反应, 即吸收热量, 因此灰分的主要成份 Si , Al , Fe 和 Ca 以及少量 Mg 等对热值也会产生影响, 同时相对 C 和 H 等非金属元素, 这些元素易电离, 不可避免会对整个等离子体特性产生一定的影响。因此, 在利用非金属加金属光谱强度建

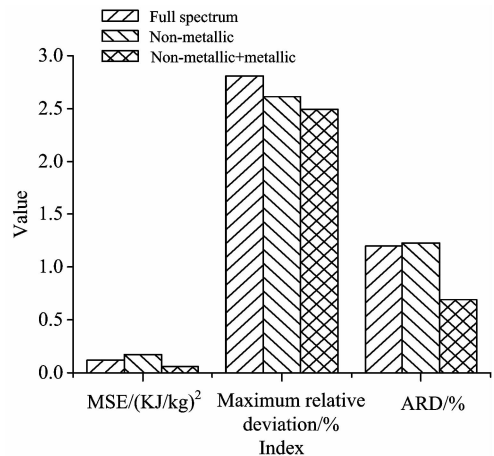


图 5 分别利用三种光谱信息预测热值评价指标对比

Fig. 5 Comparison of the evaluation index of predicting heat values using three spectral information

立的热值 SVM 回归模型预测热值时, 由于充分利用了以上对燃煤热值有影响的多种元素的光谱数据, 结果优于仅利用几种非金属元素谱线强度建立的模型。而利用全谱强度建立的 SVM 回归模型, 虽然充分纳入了以上对热值有影响的多种

元素谱线强度,但其包含的噪声、背景等谱线增加了对热值回归模型的干扰,使得模型预测可靠度有所下降,但优于只利用非金属谱线强度建模的结果。

4 结 论

将激光诱导击穿光谱应用于煤热值的定量分析,选择 44 种电厂常见的热值不同的煤样作为实验对象,提出了基于 K-CV 参数优化支持向量机算法。首先分析了燃煤热值 SVM 回归模型参数-惩罚因子 C 、核函数参数 g 对模型的影响,进而确定了 C 和 g 的最优取值空间分别为 $[-10, 10]$, $[-12,$

12]。针对燃煤 LIBS 光谱的复杂性,提出了分别将 LIBS 系统采集到的全谱强度、特征谱线(非金属谱、非金属加金属谱)强度归一化后的数据作为输入变量,用以建立 SVM 回归热值定量模型。结果表明,利用非金属加金属谱线强度建立的回归模型可以得到最佳的分析效果,其测试集 R^2 达到 0.995, MSE 为 $0.06 (\text{MJ} \cdot \text{kg}^{-1})^2$, ARD 为 0.69%。实验证明了基于 K-CV 参数优化支持向量机算法在预测煤热值的可适用性,而该预测方法结合了 K-CV 参数优化和 SVR 算法,使得定标模型具有较好的推广力,为煤质在线定量分析提供更加快速、精确的分析方法。

References

- [1] ASTM D5865—10 Standard Test Method for Gross Calorific Value of Coal and Coke. ASTM International, 2011.
- [2] ZHAO Zhong-hui, FANG Quan-guo(赵忠辉,方全国). Coal Quality Technology(煤质技术), 2017, (4): 18.
- [3] Wang Z, Yuan T B, Hou Z Y, et al. Frontiers of Physics, 2014, 9(4): 419.
- [4] Yao S C, Xu J L, Bai K J, et al. Plasma Science and Technology, 2015, 17(11): 938.
- [5] Zhang L, Ma W, Dong L, et al. Applied Spectroscopy, 2011, 65(7): 790.
- [6] Zhang L, Hu Z Y, Yin W B, et al. Frontiers of Physics, 2012, 7(6): 690.
- [7] Haider A, Rony M A, Lubna R S, et al. Optics & Laser Technology, 2011, 43(8): 1405.
- [8] DONG Mei-rong, LU Ji-dong, YAO Shun-chun, et al(董美蓉,陆继东,姚顺春,等). Journal of Engineering Thermophysics(工程热物理学报), 2012, 33(1): 175.
- [9] Haider A, Rony M A, Abedin K M. Energy & Fuels, 2013, 27(7): 3725.
- [10] Yao S, Lu J, Dong M, et al. Applied Spectroscopy, 2011, 65(10): 1197.
- [11] Yao S, Mo J, Zhao J, et al. Applied Spectroscopy, 2018, 72(8): 1225.
- [12] Lu Z, Mo J, Yao S, et al. Energy & Fuels, 2017, 31(4): 3849.
- [13] Li W, Lu J, Dong M, et al. Energy & Fuels, 2017, 32(1): 24.
- [14] Yuan T, Wang Z, Lui S L, et al. Journal of Analytical Atomic Spectrometry, 2013, 28(7): 1045.
- [15] LI Xiong-hui, WANG Zhe, LIU Han-qiang, et al(李雄威,王 哲,刘汉强,等). Infrared and Laser Engineering(红外与激光工程), 2017, 46(7): 734001.
- [16] WANG Di, LU Ji-dong, DONG Mei-rong, et al(王 帝,陆继东,董美蓉,等). Spectroscopy and Spectral Analysis(光谱学与光谱分析), 2016, 36(8): 2607.
- [17] TAN Peng, LI Xin, ZHANG Xiao-pei, et al(谭 鹏,李 鑫,张小培,等). Journal of China Coal Society(煤炭学报), 2015, 40(11): 2641.
- [18] JIANG Wen-hao, WEI Hong-qi, QU Tian-zhang, et al(江文豪,韦红旗,屈天章,等). Thermal Power Generation(热力发电), 2011, 40(3): 14.

Quantitative Analysis of LIBS Coal Heat Value Based on K-CV Parameter Optimization Support Vector Machine

DONG Mei-rong^{1,2,3}, WEI Li-ping^{1,2,3}, LU Ji-dong^{1,2,3*}, LI Wen-bing^{1,2,3}, LU Sheng-zi^{1,2,3}, HUANG Jian-wei^{1,2,3}, LI Shi-shi^{1,2,3}, LUO Fa-sheng^{1,2,3}, NIE Jia-lang¹

1. School of Electric Power, South China University of Technology, Guangzhou 510640, China

2. Guangdong Province Engineering Research Center of High Efficient and Low Pollution, Guangzhou 510640, China

3. Guangdong Province Key Laboratory of Efficient and Clean Energy Utilization, Guangzhou 510640, China

Abstract Heat value is one of the important parameters of coal quality, and it greatly affects the operation of coal-fired boilers. In order to overcome the shortcomings of traditional detection methods, laser induced breakdown spectroscopy (LIBS) was applied to quantitative analysis of heat value in coal. The structure of coal is complex. It contains many types of elements, including major, minor and trace elements, which would result the complexity of LIBS spectral information from coal. It is premise and foundation to effectively extract LIBS spectral information and achieve accurate quantitative measurement by using LIBS. In recent years, as the development of artificial intelligence technology, relevant analytical techniques have also been applied to the proximate analysis and heat value prediction of coal. In order to realize the effective extraction of LIBS spectral information in coal samples and overcome the problems such as transient fitting and poor convergence that are easily caused by conventional analytical methods, the K-fold cross validation (K-CV) parameter-optimized combined with Support Vector Machine (SVM) regression was proposed to quantitatively analyze the heat value in coal. The SVM method is an approximate realization of structural risk minimization, which can be used for pattern classification and nonlinear regression. 44 coal samples with different heat values commonly used in the power plant were selected as experimental objects, 33 of which were selected as training sets, and the remaining 11 were test sets. The correlation between the parameters of the SVM regression model-penalty factor C , the kernel function parameter g and the model accuracy were firstly analyzed based on the laser-induced coal spectrum and the best search scope for C and g were determined. Then the SVM regression model was established based on the LIBS full-spectrum and some typical elements (non-metallic elements and metal elements) feature spectra, respectively. The optimal parameters C , g of the heat value SVM regression model is obtained by using the training set spectral data, combined with the K-CV method. The spectral features of the prediction set as input are to test the reliability of the model. The calibration model established by the full spectrum, non-metallic element characteristic spectrum, as well as non-metal and metal element characteristic spectrum, respectively, could all reached 0.99, with the mean square error of 0.12, 0.17 and 0.06 ($\text{MJ} \cdot \text{kg}^{-1}$)², the forecasted average relative deviations were 1.2%, 1.23% and 0.69%. The results showed that the SVM regression method based on K-CV parameter-optimized could be used for quantitative analysis of coal heat value using LIBS technology, and could obtain higher analysis accuracy. At the same time, by comparing the quantitative analysis models using different spectral features, the quantification model of heat value by using the characteristic spectrum of non-metal plus metal elements, can effectively improve the accuracy of LIBS in the rapid detection of heat value in coal. This method can achieve accurate prediction of heat value in coal.

Keywords Coal; Heat value; Laser-induced breakdown spectroscopy; Support vector machine; Quantitative K-CV

(Received May 24, 2018; accepted Oct. 13, 2018)

* Corresponding author