

基于XY双变量特征提取策略的秸秆炭热值LIBS定量分析

段宏伟^{1,2}, 郭梅³, 朱荣光³, 牛其建^{1,2}

1. 江苏大学农业工程学院, 江苏 镇江 212013

2. 江苏大学现代农业装备与技术教育部重点实验室, 江苏 镇江 212013

3. 石河子大学机械电气工程学院, 新疆 石河子 832003

摘要 农业生物质能已逐步成为我国现代工业主要清洁能源之一。利用激光诱导击穿光谱(LIBS)技术实现秸秆炭热值(CV)精准预测。针对传统X自变量特征提取方法在LIBS定量分析秸秆炭CV过程中缺陷问题,提出了一种XY双变量特征提取法。研究首先分析了秸秆炭CV与各元素含量之间相关性,选取与CV相关性极显著($p < 0.01$)的Y型特征变量,其主要获取了以炭单质、芳香环和羧基等形式存在的C, O, H和Na元素的分析线展宽波段;同时通过筛选偏最小二乘回归(PLSR)模型回归系数阈值获取与CV相关的X型特征变量,当阈值为 4×10^{-5} 时模型交互验证均方根误差(RMSECV)降至最低值,其所对应的变量主要为参与农作物生理生长的Ca, Cr, Mg和K元素的分析线光谱线。基于所提取XY双特征变量构建遗传算法优化及自适应增强的人工神经网络(GA-BP-Adaboost)模型,当变异概率、交叉概率和相对误差率(RE)分别设为0.1, 0.95和0.01时,最优模型预测平均相对误差(AREP)和预测相对标准误差(RSDP)分别为2.39%和2.97%,相比于XY-PLSR模型效果分别降低了0.82%和0.91%。结果表明:XY双变量特征提取法结合GA-BP-Adaboost模型可以为生物质炭在工业使用过程中CV精确定量预测分析提供方法依据。

关键词 秸秆炭; 热值; 激光诱导击穿光谱; XY双变量特征提取法; 自适应增强的人工神经网络模型

中图分类号: O657.319 **文献标识码**: A **DOI**: 10.3964/j.issn.1000-0593(2022)11-3435-06

引言

随着化石能源价格不断攀升,生物质能利用价值愈发受到各国能源部门关注。生物质炭化成型燃料因清洁卫生、灰渣可回收利用等优点已成为我国主推清洁能源之一。热值(CV)是热效率、燃烧设备热平衡、燃料品质计算的重要指标,主要通过氧弹量热仪测量反应物与生成物的焓差计算得到,然而该方法难以满足工业生产实时监测要求。

激光诱导击穿光谱(LIBS)是一种利用激光在焦平面直接烧蚀样品表面产生等离子体时发射的原子或离子光谱进行检测的新兴技术^[1]。该技术在农业原料等生物质复杂基体特性精确定量分析方面仍面临挑战。Song等^[2]研究结果表明X自变量特征提取法可以去除复杂基体LIBS谱线中大量无信息变量和多重共线性变量,从而提升定量模型预测性能。

目前常规特征变量提取方法为X(光谱)自变量特征提取类型法,如竞争性自适应重加权采样法、遗传算法和回归

系数法等^[2-4],上述方法主要依据不同采样光谱波段组合构建模型的交互验证均方根误差(RMSECV)进行特征光谱提取。由于该类方法所对应回归模型主要为线性偏最小二乘回归(PLSR),而秸秆炭中主要包含金属元素原子/离子谱线和非金属元素原子/分子谱线,因此X自变量特征提取法仅能筛选出符合朗伯比尔线性定律的大部分金属元素LIBS特征光谱。

为提取与秸秆炭CV紧密相关的非金属元素LIBS特征谱线,本研究提出一种Y(浓度:热值)因变量特征提取法,主要对秸秆炭CV与元素浓度进行相关性分析,通过相关性显著分析获取大部分非金属元素最佳组合谱线。研究基于XY双变量特征提取策略精确获取秸秆炭CV敏感特征元素变量,采用遗传算法优化及自适应增强的人工神经网络算法(GA-BP-Adaboost)构建秸秆炭CV非线性LIBS定量分析模型,为秸秆炭燃料品质评价以及工业现场过程分析提供一种可靠的分析策略。

收稿日期: 2021-09-23, 修订日期: 2022-01-11

基金项目: 江苏高校优势学科建设工程(三期)项目(PAPD-2018-87), 国家自然科学基金项目(62101216), 江苏省自然科学基金项目(BK20200914)资助

作者简介: 段宏伟, 1989年生, 江苏大学农业工程学院助理研究员 e-mail: dhwsg123@ujs.edu.cn

1 实验部分

1.1 LIBS 装置

LIBS 系统烧蚀源为一台调 Q 的 Nd:YAG 固体激光器, 激发波长为 1 064 nm, 最大频率为 2 Hz, 脉冲延迟为 10 ns, 最大激发能量为 100 mJ。激光聚焦后法线入射到样品表面并引发等离子体, 相对于水平方向呈 45° 的准直镜接收等离子体信号至光纤传输通道。光纤出口连接 7 通道电荷耦合光谱仪, 光谱仪分辨率为 0.05 nm, 检测波长范围为 187~980 nm。

在本研究中, 针对激光脉冲能量波动对谱线强度影响问题, 将激光能量设为 30 mJ, 单点激光重复烧蚀次数设为 3, 单点光斑大小设为 200 μm。为了避免韧致辐射, 探测器相对于激光脉冲的延迟时间设为 0.7 μs。

1.2 样品制备

79 个秸秆炭样品采购于南京智融联科技有限公司, 包含水稻秸秆炭和玉米秸秆炭。样品经过 45 °C 烘箱干燥后, 利用刀片式粉碎机粉碎过 75 μm 筛后放入自封袋备用。样品中 N, C, S, H 和 O 采用元素分析仪测定^[5], K, Na, Ca, Mg, Pb, Cr, Cu, Zn 和 P 元素含量采用 ICP-MS 测定^[6], CV 采用氧弹量热仪测定^[7], 实验结果如表 1 所示。

表 1 CV 和主要元素含量统计结果

Table 1 Statistical results of CV and main element content

元素组成及特性	样本数	范围	均值±标准偏差
N/(g·kg ⁻¹)	79	0.63~1.41	1.01±0.22
C/(g·kg ⁻¹)	79	39.94~60.2	51.43±4.22
S/(g·kg ⁻¹)	79	0.2~0.71	0.28±0.07
H/(g·kg ⁻¹)	79	0.49~2.4	1.45±0.56
O/(g·kg ⁻¹)	79	36.23~57.15	45.83±4.56
CV/(MJ·kg ⁻¹)	79	15.68~21.17	18.73±1.3
K/(g·kg ⁻¹)	79	4.62~13.42	9.09±2.81
Na/(g·kg ⁻¹)	79	0.34~5.64	2.57±1.38
Ca/(g·kg ⁻¹)	79	2.02~10.78	5.75±2.04
Mg/(g·kg ⁻¹)	79	0.8~7.41	3.96±1.9
Pb/(mg·kg ⁻¹)	79	1.11~11.53	5.69±2.89
Cr/(mg·kg ⁻¹)	79	2.79~28.51	12.51±6.51
Cu/(mg·kg ⁻¹)	79	6.67~19.06	11.9±3.48
Zn/(mg·kg ⁻¹)	79	27.01~624.2	84.25±101.28
P/(mg·kg ⁻¹)	79	1.47~2.91	2.17±0.38

1.3 XY 双变量特征提取

定量分析模型一般形式为 $Y=KX+B$, 其中 X 和 Y 分别为自变量和因变量, 所对应特征变量提取方法可以分为 X 自变量特征提取法和 Y 因变量特征提取法。本研究所采用的 XY 双变量特征提取法流程如图 1 所示。

首先采用 Y 因变量特征提取法获取与秸秆炭 CV 相关的 C 和 O 等非金属元素谱峰波段组合, 通过 CV 与秸秆炭元素浓度之间相关性分析, 选取相关性极显著 ($p<0.01$) 元素分析线展宽波段作为特征变量。采用回归系数法 (RC) 作为 X 自变量特征提取法, 通过设定不同回归系数阈值筛选特征变

量, 并构建不同特征变量组合的线性偏最小二乘回归 (PLSR) 模型, 依据模型 RMSECV (交互验证均方根误差) 值选取最优金属元素谱线组合作为提取结果。由于所提取特征包含线性和非线性变量, 最终构建秸秆炭 CV 的非线性精确定量分析模型。

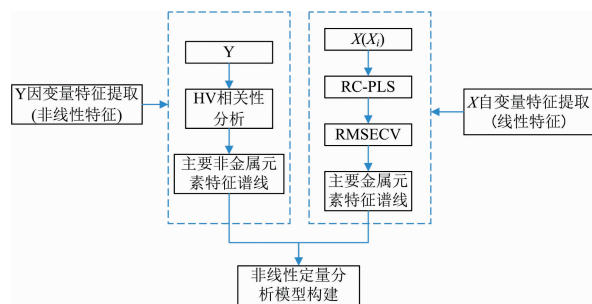


图 1 XY 双变量特征提取策略

Fig. 1 XY bivariate feature extraction strategy

1.4 模型构建及效果评价

遗传算法优化及自适应增强的人工神经网络算法 (GA-BP-Adaboost)^[8] 是一种非线性神经网络改进模型, 一方面采用遗传算法 (GA) 优化边界函数参数 w 和 b 值, 另一方面采用多个弱预测器对数据集中每个样本进行训练并不断调整样本权重 D , 同时为每个弱预测器分配权重 α 以计算最终强预测器结果。相关计算公式分别为式 (1)~式 (3)

$$\alpha_{i,j} = \frac{0.5}{\exp(\text{Error}_i)} \quad (1)$$

$$D_{i+1,j} = \frac{D_{i,j} e^{\pm \alpha_{i,j}}}{\text{Sum}(D_{i,j})} \quad (2)$$

$$\text{Error}_i = \text{Error}_i + (0, D_{i,j}) \quad (3)$$

式中, i 和 j 分别为弱预测器和样本序号, Error 为每个弱预测器所有样本权重总和。若某个样本被准确预测 (小于相对误差率阈值), 则该样本权重降低, 式 (2) 中 α 符号取负, 所对应 Error 值不变 (+0); 若某个样本未能准确预测 (大于相对误差率阈值), 则该样本权重升高, 式 (2) 中 α 符号取正, 所对应 Error 值变大 (+ $D_{i,j}$)。

模型效果主要由均方根误差 (RMSE)、平均相对误差 (AREP) 和相对标准误差 (RSD) 来评价^[9-10], 计算公式如式 (4)~式 (6) 所示。RMSECV 越小, 则建模效果越好; RSDP 和 AREP 越小, 则模型预测精度越高。

$$\text{RMSE} = \sqrt{\frac{\sum (Y_m - Y_p)^2}{n}} \quad (4)$$

$$\text{AREP} = \frac{1}{n} \sum \frac{|Y_m - Y_p|}{Y_m} \times 100\% \quad (5)$$

$$\text{RSDP} = \frac{\text{RMSEP}}{\text{mean}(Y_p)} \quad (6)$$

2 结果与讨论

2.1 LIBS 光谱分析

秸秆碳主要元素在 187~980 nm 范围内平均光谱如图 2

所示。根据 NIST 原子光谱数据库检索确认秸秆炭中含有非金属元素 C, O, H, N, S 和 P, 营养型金属元素 K, Na, Ca 和 Mg, 重金属元素 Cu, Zn, Cr 和 Pb。在农作物生长过程中, 金属元素中 K, Na, Ca, Mg, Cu, Zn 和 Pb 主要以金属阳离子形式, Cr 主要以酸根阴离子形式(CrO_4^{2-} , $Cr_2O_7^{2-}$)通过离子渗透或转运蛋白作用进入农作物各器官组织^[11]。然

而, 由于植物自身抗毒性作用, 重金属 Pb, Cu 和 Zn 大部分将被细胞壁中木栓质和果糖选择性截留^[12], 而 K, Na, Ca, Mg 和 Cr 则更容易被转运蛋白作用输送至细胞内作为植物生长功能因子。因此, 金属元素中 K, Na, Ca, Mg 和 Cr 可能会影响纤维素、半纤维素和木质素的形成, 进而影响秸秆炭 CV。

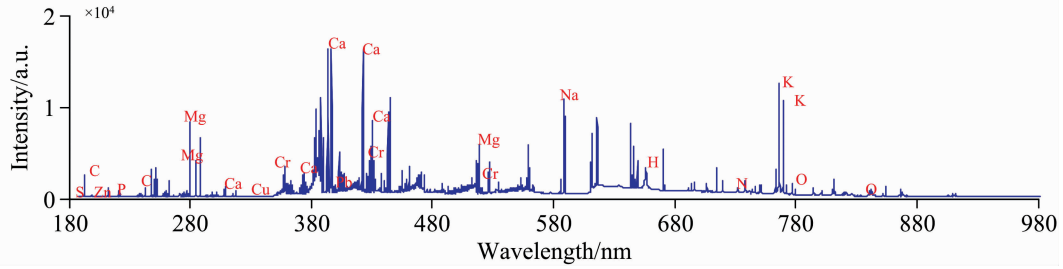


图 2 LIBS 谱线

Fig. 2 LIBS spectral analysis

与非金属元素和重金属元素相比较, 营养型金属元素 K, Ca, Na 和 Mg 发射线强度较大。原因可能是营养型金属元素的电离能较小, 更容易发生能级跃迁至激发态, 单位体积内处于激发态的原子数更多, 谱线强度更大。又由于 K, Na, Ca, Mg 和 Cr 参与农作物生理生长作用过程, 其 LIBS 分析线可能会对秸秆炭 CV 表现出更高敏感性。

2.2 XY 双变量特征提取

首先采用 Y 因变量特征提取法获取与 CV 具有显著相关性的特征元素谱峰波段, 结果如表 2 所示。CV 与 C, O, H 和 Na 元素浓度相关性较大且相关系数呈现极显著性($p < 0.01$); 与 Ca, K 和 Cr 元素浓度相关性相对较低且相关系数

呈现显著性($p < 0.05$); 而与 S, Mg, Zn, Pb, N, Cu 和 P 元素浓度无相关性。由于秸秆炭中 C, O 和 H 主要以碳单质、芳香环、羧基、醚键和硅氧键等形式存在^[13], 能够显著提升秸秆炭 CV。因此, 选取 C, O, H 和 Na 元素的 LIBS 谱峰波段作为 Y 因变量特征提取结果。

采用 X 自变量特征提取法获取全波段 PLSR 模型回归系数较大值作为 CV 特征变量, 结果如图 3(a)所示。当回归系数阈值分别设为 15×10^{-5} , 10×10^{-5} , 5×10^{-5} , 4×10^{-5} , 3×10^{-5} , 2×10^{-5} 和 1×10^{-5} 时, 其所构建 PLSR 模型的 RMSECV 结果如图 3(b)所示。结果表明: 随着阈值逐渐减小, RMSECV 值先减小后增大。原因可能是: 当 RMSECV

表 2 CV 与元素相关性分析

Table 2 Correlation analysis between CV and elements

	N	C	S	H	O	K	Na	Ca	Mg	Pb	Cr	Cu	Zn	P
CV	0.1	0.85 ^b	-0.22	0.46 ^b	-0.84 ^b	0.25 ^a	-0.33 ^b	-0.29 ^a	-0.17	-0.16	-0.22 ^a	-0.07	-0.17	0.03

注: ^a $p < 0.05$; ^b $p < 0.01$

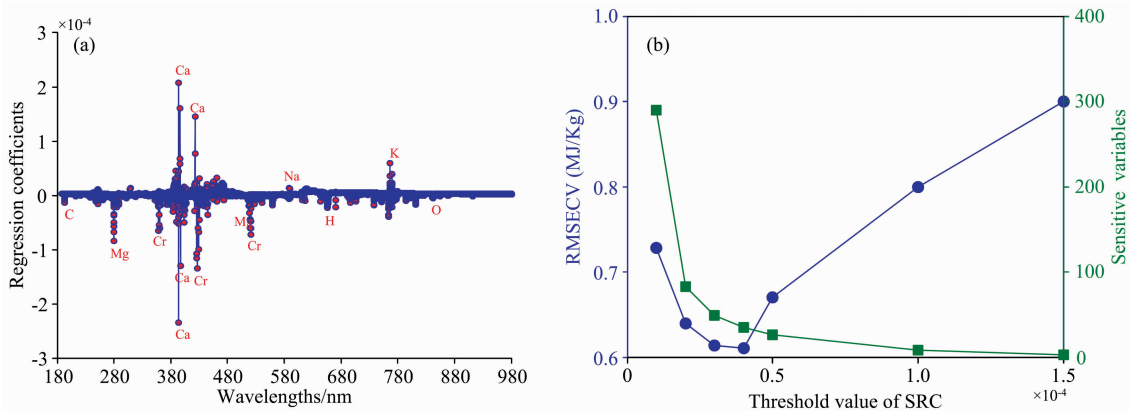


图 3 回归系数法阈值选取

(a): 全波段 PLSR 模型回归系数; (b): 不同回归系数阈值模型结果

Fig. 3 Threshold selection of regression coefficient method

(a): Full band PLSR model regression coefficient; (b): Model results with different regression coefficient thresholds

降低时, CV 模型中线性相关性元素 Ca, Cr, Mg 和 K 的分析线光谱逐渐被选中; 当 RMSECV 升高时, CV 模型中非线性相关性元素光谱和噪声信息逐渐被选中。当阈值为 4×10^{-5} 时 RMSECV 降至最小值 0.61, 所选取 49 个特征变量数为 X 自变量特征提取结果。

综上所述, XY 双变量特征提取法获取的 CV 特征变量结果如图 4(a) 所示。Y 因变量特征提取法主要包括 C, O, H

和 Na 元素分析线展宽波段, X 自变量特征提取法主要包括 Ca, Cr, Mg 和 K 元素分析线光谱, 并且二者无交集。分别构建 X 单变量、Y 单变量和 XY 双变量特征 PLSR 模型, 由图 4(b) 得出双变量特征模型的 RMSECV 值最小, 表明 XY 双变量特征提取法能够成功获取 CV 特征变量。然而由于该方法获取了大量非线性特征变量, 有必要构建非线性多元回归模型以提升模型预测性能。

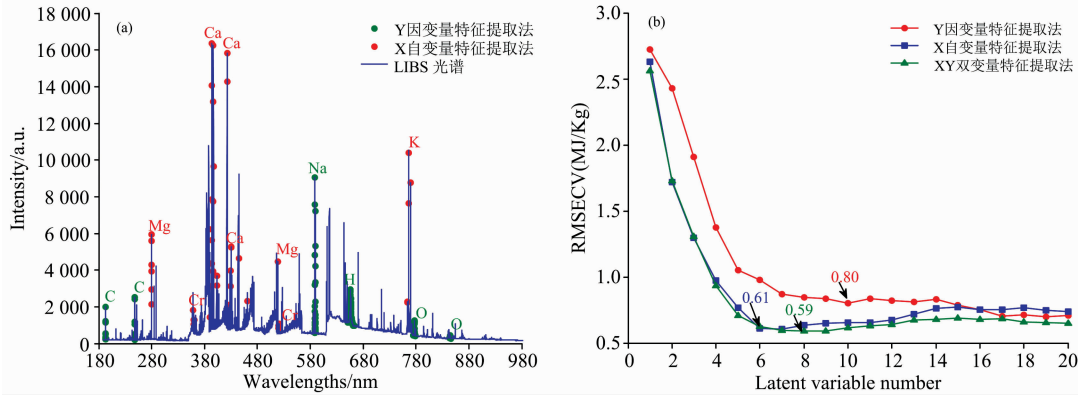


图 4 XY 双变量特征提取结果

(a): XY 双变量特征提取结果; (b): 三种特征变量模型效果比较

Fig. 4 XY bivariate feature extraction results

(a): XY bivariate feature extraction results; (b): Comparison of three feature variable model

2.3 非线性特征模型构建及预测

在构建 GA-BP-Adaboost 特征模型前, 需要提前采用 GA 和 Adaboost 对 BP-ANN 模型参数进行寻优以构建较为稳健的非线性特征模型。相关参数设置: GA 参数中迭代次数为 20, 种群规模为 20; Adaboost 参数中弱预测器个数设为 20, 双隐层数均为 1, 输入层、隐含层和输出层传递函数分别为 tansig, tansig 和 purelin; 训练函数为 trainbr, 学习速率和学习目标均设为 0.01。

对于 GA-BP 部分, 当交叉概率设为 0.8~0.95, 步长为 0.05, 变异概率设为 0.05~0.5, 步长为 0.05 时, 所对应平均适应度值如图 5(a) 所示。当变异概率逐渐减小, 交叉概率

逐渐增大时, 平均适应度值呈现降低趋势。当变异概率和交叉概率分别为 0.1 和 0.95 时, 平均适应度降至最小值 80.98。又由于适应度函数为预测绝对误差和, 适应度越小表明模型预测越准确。对于 BP-Adaboost 部分, 由于 20 个弱预测器模型对每个训练样本的相对误差率(RE)范围为 0~7.15%, 将 RE 阈值设为 0.05~0.30, 步长为 0.05, 结果如图 5(b) 所示。观察得出, 随着 RE 值增大, AREP 和 RSDP 值总体呈现先减小后增大趋势。因此, 将变异概率、交叉概率和 RE 值分别设为 0.1, 0.95 和 0.01 用于进一步的 GA-BP-Adaboost 模型构建。

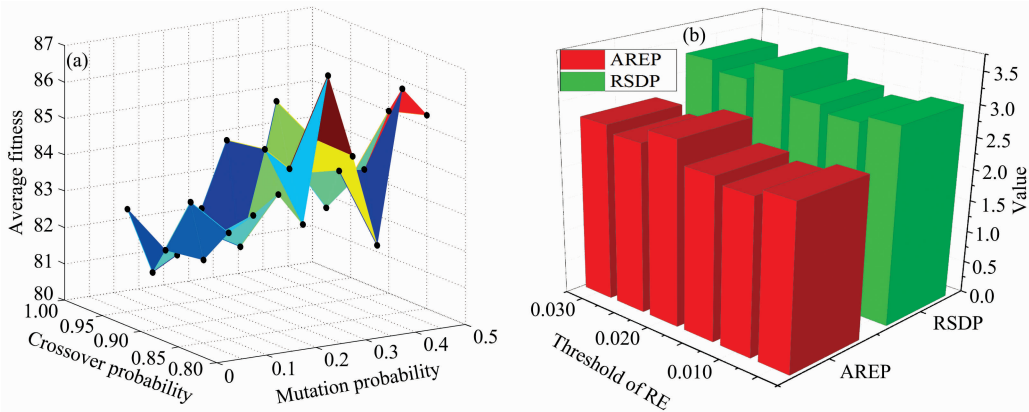


图 5 GA 和 Adaboost 参数优化

(a): GA 参数优化; (b): Adaboost 参数优化

Fig. 5 Optimization of GA and Adaboost parameters

(a): GA parameter optimization; (b): Adaboost parameter optimization

将所提取特征变量数据经主成分分析(阈值参数设为 99.93%), 并基于优化后的参数构建 GA-BP-Adaboost 模型, 将模型结果与线性 PLSR 模型效果相比较, 结果如图 6 所示。与 PLSR 特征模型相比较, GA-BP-Adaboost 特征模型中大部分样本(1~3, 5, 7~17, 19, 20)的预测值与测定值更为接近, AREP 和 RSDP 分别降低了 0.82% 和 0.91%。结果表明: 非线性 GA-BP-Adaboost 模型效果较优, 原因可能是 XY 双变量特征提取法获取了大量非线性特征变量, 而神经网络模型通过多个神经元的深度训练能够较好地实现非线性特征变量的拟合。与 Song 等^[2]所构建 CV 特征模型结果相比较, 本研究所构建 PLSR 和 GA-BP-Adaboost 模型的 RSDP 值分别下降了 0.34% 和 1.25%。原因可能是: Song 等采

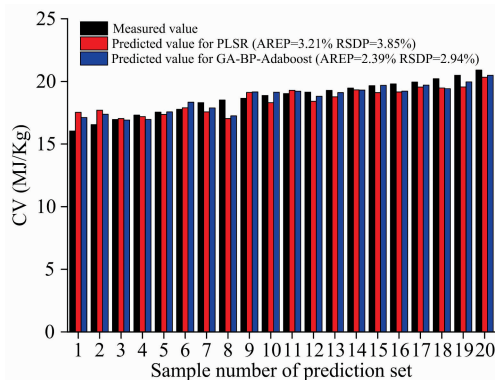


图 6 模型预测结果

Fig. 6 Model prediction results

用了一种集成变量选择法, 即将回归系数法、套索法、竞争性自适应重加权算采样法、递归加权偏最小二乘法、显著性多元相关法、最小冗余最大相关性选取的 6 个变量子集分别构建 PLSR 模型, 根据 RMSECV 值选取前 3 个变量子集并作为最终集成变量特征提取结果。然而该方法所选用的特征提取方法均属于 X 自变量特征提取法, 仅能提取符合朗伯比尔线性定律的金属元素谱线, 无法提取与 CV 相关的 C 和 O 等非金属元素特征谱线。因此, 将 XY 双变量特征提取法与 GA-BP-Adaboost 非线性多元回归模型相结合能够显著提升传统 X 自变量特征模型效果, 可以用于秸秆炭 CV 精准预测。

3 结论

针对传统 X 自变量特征提取方法在 LIBS 定量分析秸秆炭 CV 过程中缺陷问题, 提出了一种 XY 双变量特征提取法。首先分析了秸秆炭 CV 与各元素含量之间相关性, 基于相关性显著性分析结果选取与 CV 相关的 Y 型特征变量, 主要包含 C, O, H 和 Na 元素分析线展宽波段; 同时通过筛选多变量线性模型回归系数阈值获取与 CV 相关的 X 型特征变量, 主要包含 Ca, Cr, Mg 和 K 元素分析线光谱。进一步采用所提取特征变量构建秸秆炭 CV 非线性神经网络模型, 其 AREP 和 RSDP 值明显低于相关文献报道的特征模型结果。结果表明: XY 双变量特征提取法结合 GA-BP-Adaboost 模型可以用于秸秆炭 CV 精确定量预测分析。

References

- [1] Marcella D, Rim A, Alessandro D. Spectrochimica Acta Part B: Atomic Spectroscopy, 2018, 148: 105.
- [2] Song W, Hou Z, Afgan M, et al. Journal of Analytical Atomic Spectrometry, 2021, 36(1): 111.
- [3] Reda R, Saffaj T, Itqi S, et al. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 2020, 242: 118736.
- [4] Goueguel C, Soumare A, Nault C, et al. Journal of Analytical Atomic Spectrometry, 2019, 34(8): 1588.
- [5] Niu W, Han L, Xian L, et al. Energy, 2016, 100: 238.
- [6] Duan H, Ma S, Han, L, et al. Microchemical Journal, 2017, 134: 262.
- [7] Yan C, Liang J, Zhao M, et al. Analytica Chimica Acta, 2019, 1080: 35.
- [8] Wang B, Sun J, Xia L, et al. Food Reviews International, 2021, 129: 108887.
- [9] Lu P, Zhuo Z, Zhang W, et al. Applied Physics B: Lasers and Optics, 2021, 127: 82.
- [10] Yan J, Hao Z, Zhou R, et al. Analytica Chimica Acta, 2019, 1082: 30.
- [11] JIAO Xin-zhi(焦新之). Plant Physiology Journal(植物生理学通讯), 1993, 29(1): 3.
- [12] LIU Qing-quan, CHEN Ya-hua, SHEN Zhen-guo, et al(刘清泉, 陈亚华, 沈振国, 等). Plant Physiology Journal(植物生理学报), 2014, 50(5): 605.
- [13] DUAN Hong-wei, HAN Lu-jia, HUANG Guang-qun(段宏伟, 韩鲁佳, 黄光群). Transactions of the Chinese Society for Agricultural Machinery(农业机械学报), 2019, 50(10): 323.

LIBS Quantitative Analysis of Calorific Value of Straw Charcoal Based on XY Bivariate Feature Extraction Strategy

DUAN Hong-wei^{1,2}, GUO Mei³, ZHU Rong-guang³, NIU Qi-jian^{1,2}

1. School of Agricultural Engineering, Jiangsu University, Zhenjiang 212013, China

2. Key Laboratory of Modern Agricultural Equipment and Technology, Ministry of Education, Jiangsu University, Zhenjiang 212013, China

3. College of Mechanical and Electrical Engineering, Shihezi University, Shihezi 832003, China

Abstract Agricultural biomass energy has gradually become one of the main clean energy sources in modern industry in China. Laser-induced breakdown spectroscopy (LIBS) was used to accurately evaluate the calorific value (CV) of straw charcoal in this paper. Due to the defects of the traditional X -independent variable feature extraction method in LIBS quantitative analysis of CV, an XY bivariate feature extraction method is proposed. Firstly, the correlation between CV and main elemental content was analyzed, and the Y -type sensitive variables with extremely significant correlation ($p < 0.01$) were selected, which mainly obtained the analytical line broadening band of C, O, H and Na in the form of carbon, aromatic ring and a carboxyl group. Meanwhile, the X -type sensitive variables related to CV were obtained by screening the regression coefficient threshold of the partial least squares regression (PLSR) model. When the threshold is 4×10^{-5} , the cross-validation root means square error (RMSECV) decreases to the lowest value, and its corresponding variables are the analytical line spectra of Ca, Cr, Mg and K that can participate in crop growth. The GA-BP-Adaboost model was then developed based on the XY double characteristic variables extracted. When the mutation probability, crossover probability and relative error rate (RE) are set to 0.1, 0.95 and 0.01 respectively, the average relative error (AREP) and relative standard deviation (RSDP) of the optimal model are 2.39% and 2.97% respectively, which is 0.82% and 0.91% lower than that of XY -PLSR model. The results show that the XY bivariate feature extraction method combined with the GA-BP-Adaboost model can be used to accurately and quantitatively predict the CV of biomass carbon in the industrial process.

Keywords Straw charcoal; Calorific value; LIBS; XY bivariate feature extraction; GA-BP-AdaBoost

(Received Sep. 23, 2021; accepted Jan. 11, 2022)