

WCARS-ISPA 火焰光谱特征选择的转炉炼钢终点预测

朱雯琼, 周木春*, 赵琦, 廖俊

南京理工大学电子工程与光电技术学院, 江苏 南京 210094

摘要 对转炉炼钢终点的实时精准控制能够有效提高钢铁产出的质量, 炉口火焰光谱在炼钢不同时期的变化明显, 对其进行分析处理并与机器学习方法相结合可有效用于炼钢终点的实时控制。针对炉口火焰光谱数据量大、现有方法对光谱特征提取在可信度和实时性上不足的缺陷, 提出一种基于窗口竞争性自适应重加权采样(WCARS)结合迭代式连续投影算法(ISPA)的光谱特征波长选择方法, 该方法在有效解决模型过拟合问题的同时, 能够降低高维数据计算的复杂度。将火焰光谱数据沿波长方向进行窗口划分后, 使用CARS进行计算选出特征窗口波段, 再将迭代式选择与传统连续投影算法相结合, 通过重复迭代精选出特征波长, 在此基础上使用支持向量机回归(SVR)建立炼钢终点碳含量预测模型。实验采集363组炼钢后期的炉口火焰光谱数据作为样本, 并对其进行 Savitzky-Golay 平滑预处理。使用 WCARS-ISPA 算法从全光谱数据中选出10个特征波长作为 SVR 模型的输入, 碳含量为模型输出, Kennard-stone 算法对训练集和测试集进行划分, 选择碳含量的平均预测误差、预测误差在 $\pm 2\%$ 以内的命中率以及运行30次的平均时间作为模型评价指标。实验结果显示, 模型的平均碳含量预测误差为1.4132%, 命中率高达90.63%, 运行时间为0.019679 s。与使用全光谱和 WCARS-ISPA, CARS-SPA, WCARS 和 SPA 四种不同特征选择方法选出的特征波长建模得到的结果进行对比, 基于 WCARS-ISPA 方法选出的特征波长建立的终点碳含量预测模型误差最小、命中率最高。提出一种新的炉口火焰光谱特征波长提取方法, 使用窗口竞争性自适应重加权采样结合迭代式连续投影算法选取特征波长, 并在此基础上建立转炉炼钢终点碳含量预测模型, 实验结果表明, 该方法能够有效提取火焰光谱特征, 所建模型能够对转炉炼钢终点进行准确预测, 满足工业生产的实时控制要求, 为实际生产提供可靠帮助。

关键词 转炉炼钢; 火焰光谱; 窗口竞争性自适应重加权采样; 迭代式连续投影算法; 终点预测

中图分类号: TF713.4 **文献标识码:** A **DOI:** 10.3964/j.issn.1000-0593(2021)08-2332-05

引言

转炉炼钢是我国主要的炼钢技术, 它通过吹入氧气与铁水进行化学反应, 以消除铁水中的杂质, 最终产出钢铁的质量与出钢时的温度和钢铁中的成分含量密切相关, 因此对炼钢终点的精准控制尤为重要。近年来, 针对传统控制方法如人工经验控制、副枪控制、烟气分析等命中率低、成本高、难以实时控制的缺陷, 火焰光谱分析技术被提出并用于炼钢终点的控制^[1]。光谱分析早在冶金、化学等方面被广泛应用, 不仅可以用于测量温度, 也可用于检测物质成分含量, 它所提供的实时测量信息可保证生产过程的优化控制^[2]。目前, 机器学习方法随着硬件的发展, 在各个领域显示出优

越的性能, 特别在数据分析预测方面表现出色, 若将火焰光谱分析与机器学习方法相结合, 建立炼钢终点预测模型, 能够有效对炼钢过程进行实时控制, 从而提高钢铁产量、降低工业成本。但由于炉口火焰光谱数据量大, 包含大量冗余信息, 直接使用其进行建模会导致模型预测精度低且耗时长, 因此需要首先对光谱进行处理, 提取出相应特征用于建模。

目前的光谱特征提取方法主要分为两种, 一种是连续谱分段处理法, 通过对光谱整体特征的分析, 使用数学方法计算出能够代表光谱整体信息的某些参数, 如张彩军等^[3]对炉口火焰光谱进行分段最小二乘拟合得到拟合参数作为光谱稳定特征, 并对特征峰区域光强积分值作为光谱不稳定特征; Anton Stadler 等^[4]通过计算光谱的连续差异、加权差异、二值差异以及在小波域的过零点和高通差异作为光谱特征;

收稿日期: 2020-08-15, 修订日期: 2020-12-20

基金项目: 国家自然科学基金项目(61675098)资助

作者简介: 朱雯琼, 1996年生, 南京理工大学电子工程与光电技术学院硕士研究生 e-mail: 51zyxshh@163.com

* 通讯作者 e-mail: mczhou@sohu.com

SedatGolgiyaz 等^[5]通过计算功率谱密度来提取火焰光谱的闪烁特性; Chang 等^[6]对光谱进行希尔伯特黄变换(HHT),将原始数据进行经验模式分解,得到代表光谱有效信息的固有模态函数; Yin 等^[7]对火焰光谱进行傅里叶变换,并用直方图均衡和能量谱对光谱特征进行计算。另一种特征波长选取法是使用变量选择算法直接从原始光谱中选出具有显著特征的波长,如 Fan 等^[8]用竞争性自适应重加权采样(CARS)方法提取特征波长用于建立醋酸含量预测模型; Li 等^[9]用蒙特卡洛无信息变量消除方法对棉籽红外光谱进行波长选择,并将其用于测定其棉酚含量; Shao 等^[10]直接选取了光谱峰值点并进行计算处理,用于炼钢终点的分类研究。

由于光谱分段处理方法计算缓慢,无法满足工业炼钢的实时控制要求,且在计算过程中容易模糊原始光谱中的某些信息,导致预测结果差,因此常使用特征波长选择法对炉口火焰光谱进行处理,而一些传统的特征选择算法存在选取的特征可信度低、结果过拟合等问题。为解决以上方法的缺陷,提高碳含量的预测准确度,提出了一种窗口竞争性自适应重加权采样(WCARS)结合迭代式连续投影算法(ISPA)的特征变量选择方法,先用 WCARS 对原始光谱进行粗选,再用 ISPA 方法精选得到特征波长,最后使用支持向量机回归(SVR)建立炼钢终点 C 含量预测模型,并将模型预测结果与其他方法进行比较。

1 实验部分

1.1 光谱数据及预处理

数据选用了包钢炼钢厂实际冶炼的 156 个炉次共 363 组炼钢后期炉口火焰光谱数据,光谱波段为 400~1 100 nm,采样维数为 3 648。采集到的原始光谱存在较多噪声毛刺,对光谱特征的提取可能产生干扰,因此使用 Savitzky-Golay 平滑滤波对光谱进行预处理。图 1 为预处理后的炉口火焰光谱。

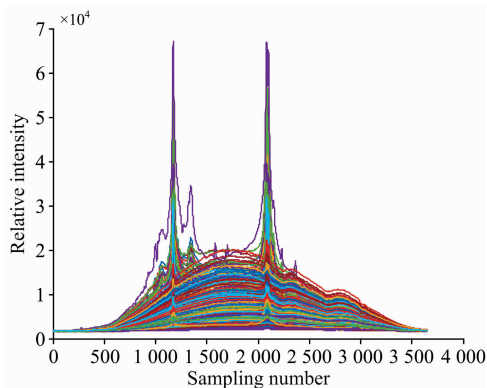


图 1 炉口火焰光谱数据集

Fig. 1 Furnace mouth flame spectrum data sets

1.2 窗口竞争性自适应重加权采样

传统竞争性自适应重加权采样过于注重校正集交叉验证结果,容易导致过拟合,而窗口竞争性自适应重加权采样(window competitive adaptive reweighted sampling,

WCARS)^[11]可以有效解决过拟合问题。WCARS 是在传统 CARS 算法的基础上,考虑相邻波长之间的协同关系,将变量沿着波长方向划分为窗口,对数据建立 PLS 模型,以窗口的回归系数绝对值的均值作为衡量窗口重要性的指标,最终利用 k 折交叉验证选出 RMSECV 最小的窗口子集。图 2 为 WCARS 的算法流程图。

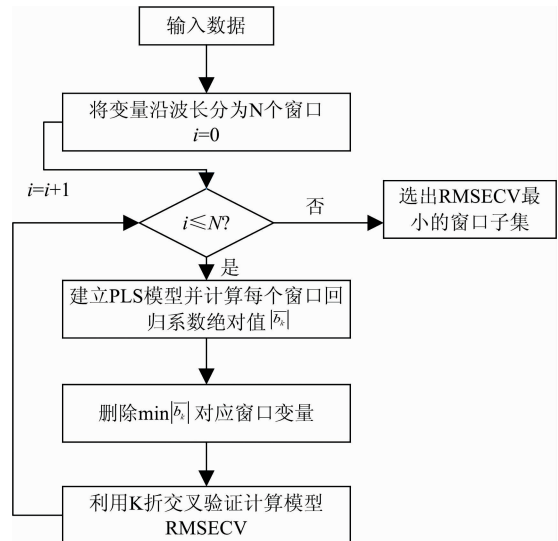


图 2 WCARS 流程图

Fig. 2 Flowchart of WCARS

1.3 迭代式连续投影算法

连续投影算法(successive projections algorithm, SPA)^[12]是一种前向变量选取法,它通过计算剩余变量与选取变量的投影向量大小来选择特征变量,能够保证选取变量间的线性关系最小,以消除变量间的冗余信息,达到选择特征波长的目的。

迭代式选择是将待选择的变量集分为 k 份,先对第一份进行特征选择,将得到的特征变量加入第二份,再对第二份进行特征选择,直到第 k 份。将迭代式选择与连续投影算法相结合,能够降低多维数据计算的复杂度,避免计算过程中的遗漏,通过重复迭代最终选出最优特征变量。

迭代式连续投影算法(ISPA)的步骤如下:

(1)将待选择的光谱矩阵按波长分为 k 份,记为 $X_{m \times n_w}$, $w=1, 2, \dots, k$, m 为光谱数, N 为需要选择的波长个数;

(2)在光谱矩阵中任选一列向量,记为 x_{j_0} ;

(3)计算剩余列向量 x_j 与当前所选向量的投影

$$P_{x_j} = x_j - [x_j^T x_{j(\rho-1)}] x_{j(\rho-1)} [x_{j(\rho-1)}^T x_{j(\rho-1)}]^{-1}, \quad \rho = 1, 2, \dots, N \quad (1)$$

(4)取投影向量最大的变量序号

$$j_\rho = \arg[\max(\|P_{x_j}\|)] \quad (2)$$

(5)令 $x_j = P_{x_j}$, $P = P + 1$, 返回第(3)步,直到 $\rho = N$;

(6)将得到的 N 个特征变量加入下一份待选择变量中,返回第(2)步进行计算,直到 $w = k$ 。

1.4 支持向量机回归

支持向量机回归(support vector regression, SVR)是根

据结构风险最小化原则,以寻找一个最优超平面为目的机器学习算法。SVR 通过核函数将原始数据映射到高维空间,保证数据集可以与目标值具有最大偏差 ϵ ,同时平坦度应尽可能高^[13]。应用线性学习机的方法解决样本空间中的非线性问题,具有很强的非线性建模能力。本文使用 SVR 建立碳含量预测模型,选用径向基函数(RBF)作为核函数,使用粒子群算法选择最佳参数 c 和 g ^[14]。

2 结果与讨论

2.1 WCARS 计算结果

在进行波长筛选之前需要先设定合适的窗口数目对光谱进行划分。一帧炉口火焰光谱有 3 648 个波长采样点,窗口太少可能导致有效数据与大量冗余数据被分在同一个窗口,降低了被选中的概率,太多则与传统 CARS 无异,因此选择 150~450 个窗口计算建模,根据模型 RMSE 选择最佳窗口数。

计算得到的 RMSE 随选择窗口数的变化如图 3 所示。可以看出随着窗口数的增大 RMSE 先逐渐减小,在窗口数为 400 时达到最小,而后重新增大。因此选择 400 个窗口对原始光谱进行划分。图 4 为当窗口数为 400 时,WCARS 对原始光谱计算得到的结果,共选出 48 段特征波长子集,包含 434 个波长。

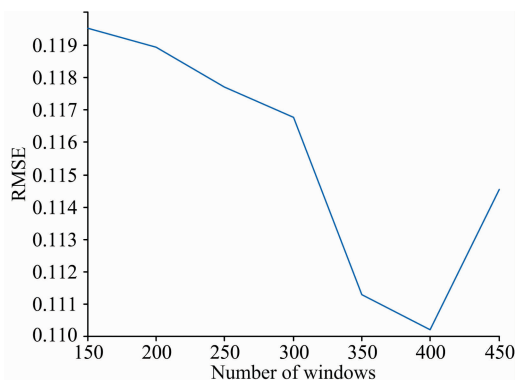


图 3 RMSE 与窗口数的变化情况

Fig. 3 Variation of RMSE with the number of windows

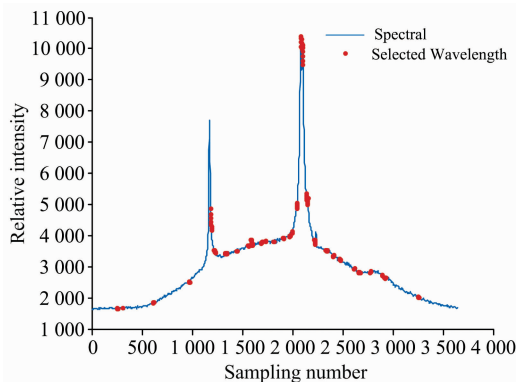


图 4 WCARS 粗选结果

Fig. 4 Rough selection result of WCARS

2.2 ISPA 计算结果

WCARS 以窗口为单位进行波长选择,每个窗口中包含多个相邻波长,其互相影响会导致一些干扰信息的存在,且粗选得到 434 个波长,数据量仍较大,影响模型预测速度。因此使用 ISPA 算法对选出的波长进行精选,通过消除变量间的无效冗余信息来进一步压缩数据。

图 5(a)为用 ISPA 对粗选波长进行计算得到的结果,可以看出随选择变量数的增多,模型 RMSE 迅速减小,当选择变量数为 10 时, RMSE 达到最小值,为 0.094 482,而后增大。图 5(b)为传统 SPA 对粗选波长的计算结果,当选择变量数为 8 时, RMSE 最小为 0.096 096,可以看出 ISPA 相比 SPA 结果更好,且 ISPA 所选波长包含这 8 个波长。因此选用 ISPA 计算得到的 10 个特征变量作为最终结果,这 10 个特征波长序号为 252, 253, 971, 1 189, 1 217, 1 502, 2 039, 2 142, 2 662 和 2 782,如图 6 所示。

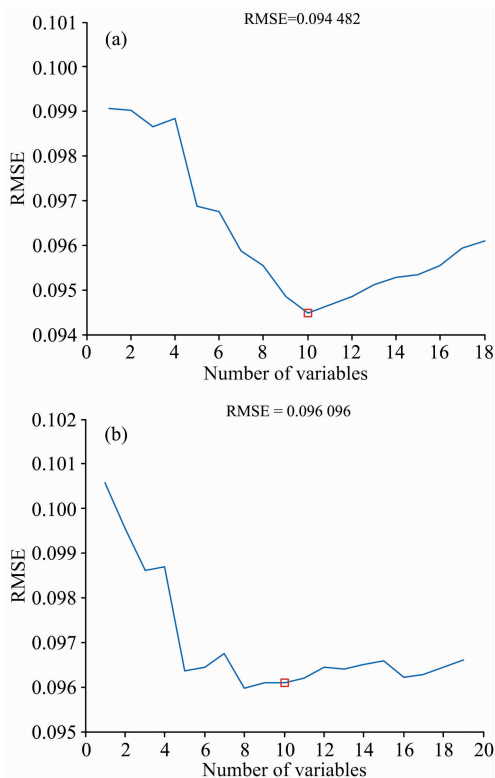


图 5 RMSE 与选择变量数的变化情况

(a): ISPA; (b): SPA

Fig. 5 Variation of RMSE with the number of selected variables

(a): ISPA; (b): SPA

2.3 基于 WCARS-ISPA 的终点碳预测模型结果

炼钢过程十分复杂,光谱特征与碳含量的关系并非简单的线性关系,支持向量机回归(SVR)对于非线性建模问题具有很强的能力,因此常用于炼钢终点模型的建立。SVR 是一种有监督的学习方法,其性能取决于训练和测试数据集,模型的输入参数对最终结果的影响极大,因此,光谱特征的选取对最终碳含量的预测十分重要。为验证 WCARS-ISPA 算

法选取出的火焰光谱特征波长点用于预测炼钢终点碳含量的有效性, 本文使用 SVR 建立终点碳含量预测模型。

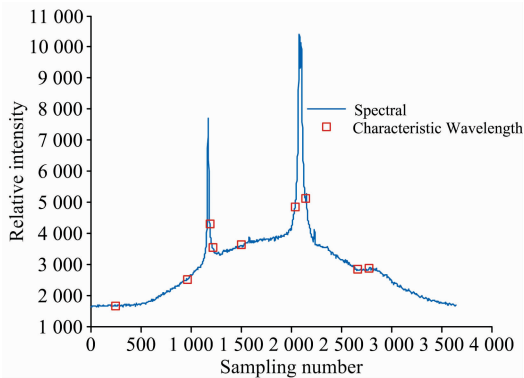


图 6 WCARS-ISP 特征波长选择结果

Fig. 6 Selection result of characteristic wavelengths with WCARS-ISP

将选取的 10 个光谱特征波长与 16 个炉口火焰图像特征、2 个峰值特征共 28 个变量作为输入, 碳含量为输出。使用 kennard-stone 算法对训练集和测试集进行划分, 将 363 个数据分为 290 个训练数据和 73 个测试数据。将 C 含量的平均误差、预测误差在 $\pm 2\%$ 以内的命中率以及运行 30 次的平均时间作为模型评价指标。图 7 为 WCARS-ISP-SVM 模型的训练结果, 从图中可以看出预测 C 含量与实际数据基本重

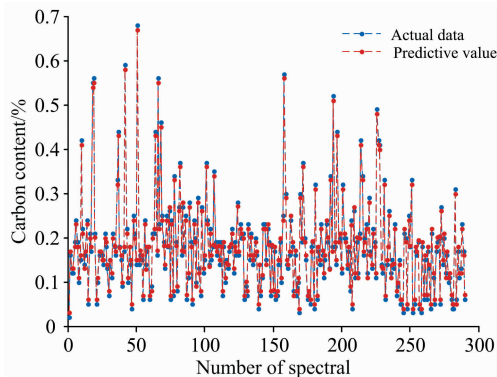


图 7 WCARS-ISP 模型的训练结果

Fig. 7 Training results of the WCARS-ISP model

合。WCARS-ISP-SVR 模型得到的平均 C 含量误差为 1.413 2%, 命中率高达 90.63%, 平均时间为 0.019 679 s, 能够满足工业生产实时预报要求。

为证明本方法的优越性, 分别使用全光谱和 WCARS-ISP, CARS-SPA, WCARS, SPA 四种不同特征选取方法选出的特征波长建模, 并对其结果进行比较。表 1 总结了五个模型的三项评价指标, 从表中可以看到, 使用全光谱建立终点碳预测模型得到的平均预测误差高达 3.369 1, 远远超过其他模型, 命中率低且运行时间过长, 无法满足工业生产要求, 因此说明对原始光谱进行特征提取是必要的。而与其他三种常用特征选取方法相比, WCARS-ISP 模型得到的终点碳平均预测误差更小, 且命中率最高, 说明该方法效果更好, 能够进一步提高炼钢终点碳含量的预测精度。

表 1 不同模型的预测结果

Table 1 Prediction results of different models

模型	特征波长数	平均误差/%	命中率/%	运行时间/s
WCARS-ISP	10	1.413 2	90.63	0.019 679
全光谱	3 648	3.369 1	77.13	0.822 247
WCARS	434	2.135 0	82.37	0.103 025
SPA	10	1.881 5	84.02	0.017 702
CARS-SPA	8	1.820 9	86.77	0.017 809

3 结 论

针对转炉火焰光谱数据量大、传统特征选择算法选取的特征可信度低、结果过拟合等问题, 提出了一种 WCARS-ISP 算法, 对炉口火焰光谱进行特征波长选取, 并在此基础上使用 SVR 建立炼钢终点碳含量预测模型。先使用 WCARS 对炉口火焰光谱进行粗选, 再用 ISP 对选出波长进一步精选, 最终选出 10 个特征波长作为 SVR 输入变量。该方法能够得到较好的实验结果, 模型预测平均碳含量误差为 1.413 2%, 误差在 $\pm 2\%$ 以内的命中率高达 90.63%, 运行时间小于 0.02 s, 结果优于现有其他波长选择方法。将该模型用于实际生产, 能够有效地对转炉炼钢终点进行控制, 满足炼钢终点实时预测的需求, 帮助降低工业成本、提高钢铁产量。

References

- [1] FENG Shi-chao, WANG Yan-hong, DING Rui-feng(冯士超, 王艳红, 丁瑞锋). Metallurgical Industry Automation(冶金自动化), 2016, 237(2): 1.
- [2] CHU Xiao-li, SHI Yun-ying, CHEN Pu, et al(褚小立, 史云颖, 陈 瀑, 等). Journal of Instrumental Analysis(分析测试学报), 2019, 38(5): 603.
- [3] ZHANG Cai-jun, HAN Yang, HE Shi-yu, et al(张彩军, 韩 阳, 何世宇, 等). Chinese Journal of Scientific Instrument(仪器仪表学报), 2018, 39(1): 24.
- [4] Stadler A, Windisch T, Diepold K. Fire Safety Journal, 2014, 66(5): 1.
- [5] Golgiyaz S, Talu M F, Onat C. Fuel, 2019, 255: 115827.
- [6] Chang J C, Wang J C, Wang Z, et al. Complexity, 2018, <https://doi.org/10.1155/2018/8682725>.

- [7] Yin Z J, Luo Q, Tang T T, et al. *Advances in Engineering Research*, 2016, 116: 341.
- [8] Fan W, Shan Y, Li G Y, et al. *Food Analytical Methods*, 2012, 5(3): 585.
- [9] Cheng L, Zhao T L, Li C, et al. *Food Chemistry*, 2017, 221(Apr. 15): 990.
- [10] Shao Y L, Zhao Q, Chen Y R, et al. *Optik*, 2015, 126(23): 4539.
- [11] LI Pao, ZHOU Jun, JIANG Li-wen, et al(李 跑, 周 骏, 蒋立文, 等). *Spectroscopy and Spectral Analysis(光谱学与光谱分析)*, 2019, 39(5): 1428.
- [12] Tang R N, Chen X P, Li C. *Applied Spectroscopy*, 2018, 72(5): 740.
- [13] Zhang F, O'Donnel Lauren J. *Support Vector Regression*. in: *Machine Learning*, Academic Press, 2020, 123.
- [14] Tang B W, Zhu Z X, Shin H S, et al. *Information Sciences*, 2017, 420: 364.

End-Point Prediction of BOF Steelmaking Based on Flame Spectral Feature Selection Using WCARS-ISPA

ZHU Wen-qiong, ZHOU Mu-chun*, ZHAO Qi, LIAO Jun

School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing 210094, China

Abstract Real-time precise control of the BOF steelmaking end-point can effectively improve the quality of steel output. The flame spectra change obviously in different stages of steelmaking. It can be used to control the end-point of steelmaking effectively with the machine learning method. Due to a large amount of spectral data and the lack of reliability and real-time performance of the existing methods for spectral feature extraction, a characteristic spectral wavelength selection method based on window competitive adaptive reweighted sampling (WCARS) combined with iterative successive projection algorithm (ISPA) was proposed in this paper. This method can effectively solve the problem of over-fitting and reduce the complexity of high-dimensional data calculation. After dividing the spectral data along the wavelength direction in the window, CARS was used to select the feature window band. The iterative selection was combined with a traditional successive projection algorithm, and the characteristic wavelengths were selected through repeated iteration. On this basis, support vector machine regression (SVR) was used to establish the carbon content prediction model of steelmaking end-point. 363 sets of spectral data of the later stage of steelmaking were collected as an experimental sample and preprocessed by Savitzky-Golay smoothing. The input of the SVR model was 10 characteristic wavelength data selected by WCARS-ISPA, and the output was carbon content. The training set and test set were divided by the Kennard-Stone algorithm. The average prediction error of carbon content, the hit ratio of prediction error within $\pm 2\%$ and the average running time of 30 times were selected as the evaluation indexes. The results indicated that the average prediction error is 1.413 2%, the hit ratio is 90.63%, and the running time is 0.019 679 s. Compared with the model of full spectra and characteristic wavelengths selected by four different feature selection methods of WCARS-ISPA, CARS-SPA, WCARS and SPA, the WCARS-ISPA model has the lowest error and the highest hit ratio. In this paper, a new flame spectral characteristic wavelength extraction method was proposed. Window competitive adaptive reweighted sampling was combined with an iterative successive projection algorithm to select the wavelength, and a prediction model of end-point carbon content is established on this basis. The experimental results showed that this method could effectively extract the spectral characteristics of flame. This model can accurately predict the endpoint of BOF steelmaking and meet the requirements of real-time control of industrial production.

Keywords BOF steelmaking; Flame spectra; Window competitive adaptive reweighted sampling; Iterative successive projection algorithm; End-point prediction

(Received Aug. 15, 2020; accepted Dec. 20, 2020)

* Corresponding author