

LS-SVM 算法在近红外光谱煤质分析中的应用^①

王 帅, 张云佳, 王雪梅, 乔建仙

(解放军后勤工程学院 化学与材料工程系, 重庆 401331)

摘 要: 针对目前电力行业煤质分析的需求, 提出了基于 Hadamard 近红外光谱的煤质分析技术, 对 Hadamard 近红外光谱仪研制、控制分析软件设计、煤炭光谱信号采集、指标特征信息提取、定量模型建立五个环节综合考虑, 研发了 Hadamard 近红外煤质分析系统. 研究中, 对 41 个不同质量指标的标准煤样进行了定量分析预测, 考察了在相同粒径的条件下 Hadamard 近红外光谱对煤炭指标的预测能力, 提出了基于 ICA+LS-SVM 算法的煤炭指标预测方法, 光谱数据与煤炭指标具有很好的相关性, 相关系数普遍在 0.9 以上, 取得了较好预测效果.

关键词: 近红外光谱; 煤质分析; Hadamard 变换; 最小二乘支持向量机; 快速检测

Application of LS-SVM Algorithm in Coal Quality Indicators Prediction Using Near-Infrared Spectroscopy

WANG Shuai, ZHANG Yun-Jia, WANG Xue-Mei, QIAO Jian-Xian

(Department of Chemistry and Material Engineering, Logistic Engineering University, Chongqing 401331, China)

Abstract: In order to meet the demand of coal quality analysis in power industry, this paper proposes a new technology of coal analysis based on Hadamard near infrared spectroscopy. On the basis of comprehensive consideration of five aspects, such as development of near infrared spectrometer, the design of control and analysis software, the acquisition of coal spectral signal, the extraction of characteristic information and the establishment of quantitative model, the Hadamard near infrared coal quality analysis system is developed. In the study, 41 standard coal samples with different quality indicators are quantitatively analyzed and predicted. The ability of Hadamard near infrared spectroscopy to predict coal indicators in conditions of same particle size is investigated. Based on LS-SVM algorithm, a coal indicators prediction method is proposed, which has a good correlation with the coal index, and the correlation coefficients is above 0.9.

Key words: near infrared spectroscopy; coal analysis; Hadamard transform; ls-svm algorithm; rapid detection

煤炭作为我国的主要能源, 对我国国民经济的发展起着极其重要的作用. 随着经济的发展和人民生活水平的不断提高, 人们对于煤炭的利用和研究也越来越广泛和深入. 火力发电在我国电力生产行业中占主导地位, 月发电量达到 4082.60 亿千瓦时, 占总发电量的 83.60%^[1,2]. 而煤电企业主要原料是煤炭, 占到了公司总成本的一半以上, 长期以来, 火力发电厂燃煤煤质成分的分析一直都是采用实验室分析的手段, 分析一个煤样需要六个小时以上, 远远不能满足燃烧调整的需要, 因此, 实现煤质的快速分析已成为工业生产实践的迫切需要^[3]. 对于火力发电厂而言, 燃煤费用约

占其运营成本的 70%, 因此入场煤与入炉煤的煤质监理工作显得十分的重要, 直接关系到电厂的安全生产与运行. 全国煤炭市场开放以后, 电厂燃煤中, 由国家统一配送的比例不断下降, 很多小煤窑成为了电厂燃煤供给的主体, 由此带来了一系列的问题, 其中煤质多变这一问题显得尤为突出^[4-6]. 传统的煤质分析方法一般是基于化学方法, 煤样由人工采集制备, 分析速度缓慢, 难以满足电力企业的需要, 一批煤样在进入分析流程后, 需要等待 6 至 8 小时才能得到分析结果, 当天无法得到煤质检测报告, 这样的检测效率无法满足发电厂锅炉调整的需要. 由于传统检测方式方

① 基金项目:2015 重庆市基础与前沿研究计划(cstc2015jcyjA100009);2014 解放军后勤工程学院青年基金

收稿时间:2016-08-11;收到修改稿时间:2016-09-18 [doi: 10.15888/j.cnki.csa.005726]

法的局限,使得煤质快速分析检测需求与检测数据报出时间滞后的矛盾日益突出,所以很有必要研发一种煤质快速在线监测分析的新方法^[7,8]。

近年来,如何更加合理、洁净、有效地利用煤炭,是目前煤化学研究领域中所面临的一个重要问题。由于近红外光谱分析技术具有效率高、成本低等自身的优点而备受煤化学工作者的青睐。本文针对目前电力行业煤质分析的需求,提出了基于Hadamard近红外光谱的煤质分析技术,研发了Hadamard近红外煤质分析系统,并进一步考察了在相同粒径的条件下Hadamard近红外光谱对煤炭指标的预测能力,提出了基于ICA+LS-SVM算法的煤炭指标预测方法。

1 Hadamard近红外光谱仪硬件架构

煤炭属于固体样品,因此需采用漫反射红外光谱技术进行检测。尽管漫反射近红外光谱在烟草、制药以及农业领域有广泛的应用,但是在煤质分析这一领域的应用却较少,究其原因主要有以下三个方面:(1)煤炭属于深色样品,对于光线的反射较弱,因此分析光携带的特征信息量较少,为后期数据挖掘带来较大难度。相比较而言,农产品、药材的颜色较浅,更适用漫反射光谱技术。(2)煤炭的均匀性差,光谱数据离散性较大,不同样品区域的表面性质不同,反射光不同,光谱数据需要经过反复采集和平均处理,因此谱线的位置波动较大,这也为光谱数据特征的提取增加了难度。(3)从仪器实现角度而言,煤质分析检测不同于普通的农产品检测,由于火电厂的工况环境较差,要求仪器具备良好的抗振性和防潮性,仪器的光路设计中要尽可能避免引入精密可动光学部件,目前常用的分光技术在这一点上都受到不同程度的制约。而Hadamard数字分光技术的实现方式是DMD数字微镜芯片,由于是电路设计并采用了芯片封装,因此整个分光核心具备了很高的抗振性和稳定性,可以满足目前煤质分析的需要。

Hadamard近红外煤质分析系统主要可以分为以下几个模块:

(1) 光路模块:产生近红外区域连续波长的复色光,并使用一系列的组合镜对复色光进行聚焦,以提高入射光强度,从而适应深色物质的检测;

(2) 机械模块:为光谱仪器提供金属外框架支撑,同时提供仪器系统运转的动力,主要使用步进电机实

现仪器系统各个部件的精确定位;

(3) 电路控制模块:主要实现对光路模块、机械模块的自动化控制,同时在硬件电路底层实现简易、基本的检测信号数学变换,获取原始的光谱检测数据。

以上三大模块相互关联、相互影响,不能绝对隔离开来,只是为了便于对于整个仪器系统的描述,所以才将Hadamard近红外煤质分析系统划分为三个模块,具体划分如图1所示。

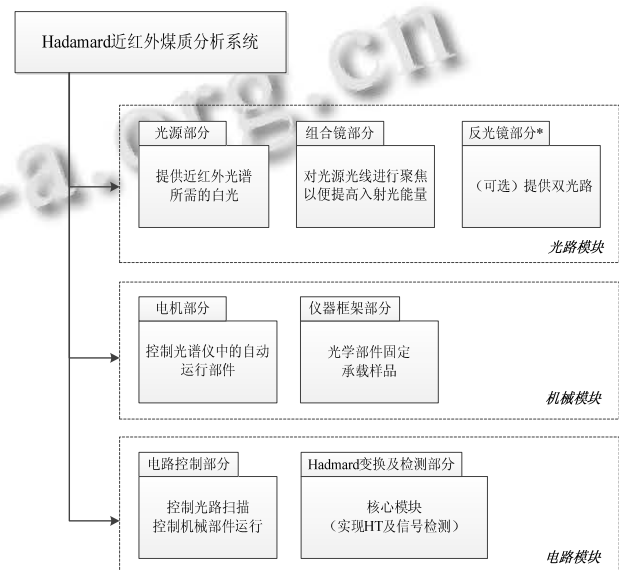


图1 Hadamard近红外煤质分析系统的模块划分

Hadamard近红外煤质分析系统的控制采用“PC+下位机”的控制方式实现,一方面便于仪器系统的快速研发与搭建,另一方面也便于系统的持续改进与完善,同时光谱数据分析、算法设计也存在大量的未知因素,需要根据光谱数据的特征进行定制开发,因此使用PC控制方式便于引入复杂分析算法,也让便于光谱数据检测、分析、预测的一体化集成。

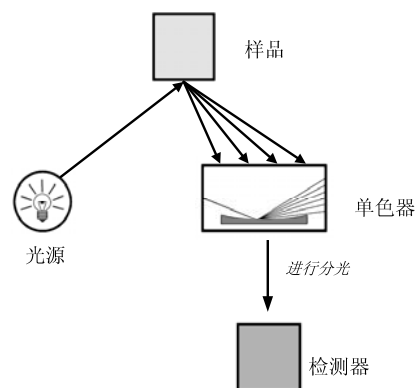


图2 反射近红外光谱仪光路示意图(后分光型)

仪器的光路采用“斜射反射式”的后分光光路,其原理图如图2所示,光源发射的复色光以一定角度斜射样品,而后携带特征信息的反射光进入单色器,进行分光数学变换,最终由检测器转换为电信号,经A/D电路转换得到数字光谱.仪器的具体光路设计见图3.

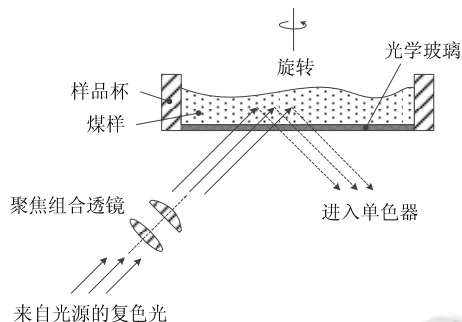


图3 改进后的煤粉检测光路



图4 HNID-2G分析平台原型机实物图

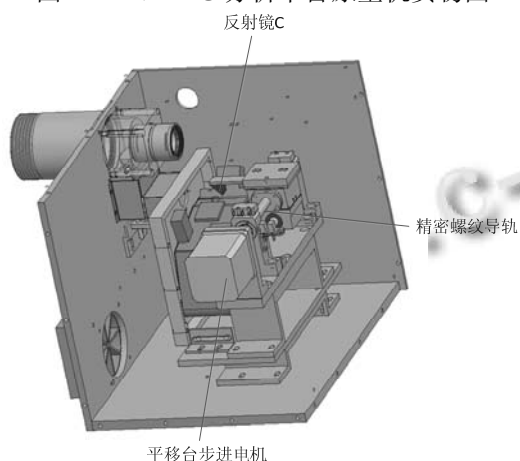


图5 光谱仪内部机械结构

2 试验样品

本组试验样品为标准煤样,所有指标按照国标方法进行标定,所有样品的粒径均为80目.主要用于考察在煤粉粒径相同的前提下,Hadamard近红外光谱对

不同煤样的指标预测能力.

试验样品集:标准煤样41个,从国家煤炭质量监督检验中心购买16个,从北京纳克分析仪器分析公司购买25个.样品包含烟煤和无烟煤,指标覆盖范围较广,具有一定的代表性.样品图见图6,样品指标见表1(样品的工业分析指标)、表2(样品的元素分析指标).



图6 样品集煤样

表1 样品集的工业分析指标

编号	样品编码	灰分 (%)	挥发分 (%)	热值 (MJ/kg)	真空相对密度 20°C	备注
1	GBW11101m	9.64	23.94	32.08	1.38	烟煤
2	GBW11101n	35.58	23.23	21.15	1.65	烟煤
3	GBW11102j	23.73	30.08	23.76	1.58	烟煤
4	GBW11102k	31.05	25.58	21.33	1.65	烟煤
5	GBW11103f	16.30	6.57	26.34	1.95	无烟煤
6	GBW11103g	24.38	5.39	23.93	2.04	无烟煤
7	GBW11104f_a	13.80	7.02	29.43	1.65	无烟煤
8	GBW11105d	11.45	8.23	31.21	1.48	无烟煤
.....
40	GBW(E)110045	8.62	33.44	30.75	1.39	烟煤
41	GBW(E)110046	10.45	17.65	31.57	1.42	烟煤

表2 样品集的元素分析指标

编号	样品编码	全硫 (%)	碳 (%)	氢 (%)	氮 (%)	备注
1	GBW11101m	0.55	79.27	4.40	1.31	烟煤
2	GBW11101n	0.47	52.61	3.45	1.00	烟煤
3	GBW11102j	1.61	59.95	3.63	1.06	烟煤
4	GBW11102k	1.45	53.42	3.50	0.97	烟煤
5	GBW11103f	0.28	77.84	0.90	0.23	无烟煤
6	GBW11103g	0.39	70.95	0.76	0.30	无烟煤
7	GBW11104f_a	1.13	79.13	2.23	1.13	无烟煤
8	GBW11105d	1.61	80.30	3.18	1.07	无烟煤
.....
40	GBW(E)110045	1.05	75.40	4.76	1.34	烟煤
41	GBW(E)110046	0.47	79.32	4.04	1.30	烟煤

3 基于局部最小二乘支持向量机定量分析 (LS-SVM)

虽然人工神经网络由于对非线性函数可任意逼近而在红外光谱分析中被广泛使用,但由于其泛化能力较弱,存在过拟合问题,且由于近红外光谱仪测得的样本光谱数据通常有几百甚至上千维,用人工神经网络方法建立模型,必须对原始光谱数据采用特征提取等降维技术来减小计算量,在实际的应用中受到一定的限制^[9,10].

支持向量机(Support Vector Machine, SVM)是一种新的机器学习方法,其基础是 Vapnik 创建的统计学习理论(Statistical Learning Theory, SLT). 统计学习理论采用结构风险最小化(Structural Risk Minimization, SRM)准则,在最小化样本点误差的同时,最小化结构风险,提高了模型的泛化能力,且没有数据维数的限制,可解决人工神经网络在煤炭漫反射光谱分析中遇到的问题. 由于 SVM 可进行函数回归,因此可应用于煤炭的漫反射光谱定量分析中.

LSSVM 作为 SVM 的一种改进技术,在求解过程中将二次规划问题转化为一组线性方程的求解,使得计算速度有很大提高^[11,12]. 在通常的光谱分析中,这些应用均取得较好结果,其模型精度好于 PLS 等线性方法. 然而, SVM 参数的确定与样本间的距离有关,通常其优化是全局优化,对于局部样本,未必能得到最优的回归结果,有时反而会造成个别点误差较大. 为此,研究中使用了一种基于局部 LSSVM 的光谱分析方法. 这种方法克服了局部非线性造成的影响,针对每次建模训练样本的不同,自动优化 LSSVM 的参数,使得预测精度有较大的提高.

1) LSSVM 的基本原理

LSSVM 用于函数拟合的原理叫简单描述如下: 给定一个有 N 个样本数据组成的训练集 $\{x_k, y_k\}_{k=1}^N$, 其中输入数据 $x_k \in R^m$, 输出数据 $y_k \in R$. 函数拟合问题可以描述为以下最优化问题:

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

$$s. t. \quad y_k = w^T \phi(x_k) + b + e_k \quad (k=1, \dots, N)$$

其中 $\phi(\cdot): R^m \rightarrow R^{m_h}$ 用于将输入数据从空间 R^m 映射到高维特征空间 R^{m_h} , $w \in R^{m_h}$ 为加权向量, $\gamma > 0$ 为惩罚系数,可以调整误差所起的作用, $e_k \in R$ 为误差变量, $b \in R$ 偏置值. 式(1)的最优化问题可变换到对

偶空间加以求解, 可得 LSSVM 拟合模型为:

$$y(x) = \sum_{k=1}^N \alpha_k \cdot K(x_k, x) + b \quad (2)$$

其中 α_k 表示支持向量, α_k 和 b 可根据训练样本数据求得. $K(x_i, x_j)$ 称为核函数, 通常核函数采用高斯核函数形式:

$$K(x_i, x_j) = \exp(-\|x_j - x_i\|^2 / \sigma^2) \quad (3)$$

2) 局部 LSSVM 的算法实现

采用高斯核函数的支持向量机, $\|x_j - x_i\|^2$ 表示样本 x_j 到 x_i 欧氏距离的平方. 因此, 其预测结果取决于该测试样本与训练样本间的欧式距离. 而参数 σ^2 用于调整欧氏距离对预测结果影响作用的大小, 通常 σ^2 的确定由训练样本交叉检验的均方误差 (RMSECV) 决定^[11]. 对于不同的训练样本, σ^2 的取值范围可能会相差很大. 如果用全部样本来确定 σ^2 , 得到的参数只是全局最优, 这时采用局部建模, 该参数对于某一试样本而言未必是最优值. 而如果对每次局部建模的样本都进行优化, 计算时间和复杂程度均很大. 因此, 需要一种自动优化模型参数的方法.

现定义测试样本和训练样本的平均距离 AD 为:

$$AD = \frac{1}{N} \sum_{i=1}^N \|x_{test} - x_{train}(i)\|^2 \quad (4)$$

其中 x_{test} 表示测试样本, $x_{train}(i)$ 表示第 i 个训练样本.

由于高斯核函数对样本的预测由距离决定, 同时用 σ^2 来调整距离的作用大小. 因此, 可以用相对距离

$RD = \frac{\|x_j - x_i\|^2}{AD}$ 来改造高斯核函数, 式(3)转换为:

$$K(x_i, x_j) = \exp\left(-\frac{\|x_j - x_i\|^2}{AD \cdot Gain}\right) \quad (5)$$

其中 $Gain$ 为相对距离增益.

局部 LSSVM 算法采用式(5)高斯核函数. 首先对某一测试样本根据欧式距离选取最近的 i 个训练样本作为建模样本集, 求得平均距离 AD , 最后进行 LSSVM 建模, 并实现对训练样本的预测. 其核心是减少训练样本集的数量, 对测试样本影响微小的训练样本不参与建模, 以提高建模精度.

4 基于LS-SVM算法的指标预测

使用的 ICA+LS-SVM 支持向量机算法对煤炭指标进行预测, 将原始光谱信号进行 ICA 分析, 得到混

合矩阵, 然后使用该混合矩阵代替原始光谱数据, 使用 LS-SVM 算法进行指标预测. 算法的具体思路见图 7, 图中测试数据 1 和测试数据 2 是煤炭样品的两组红外光谱实验数据, 针对每一组测试数据利用 LS-SVM 算法预测煤炭指标, 得到相应的指标预测结果. 将指标预测 1 与指标预测 2 进行比较, 考查算法的重复性, 将指标预测 1、指标预测 2 与煤炭样品指标真实值进行比较, 考查算法的准确性.

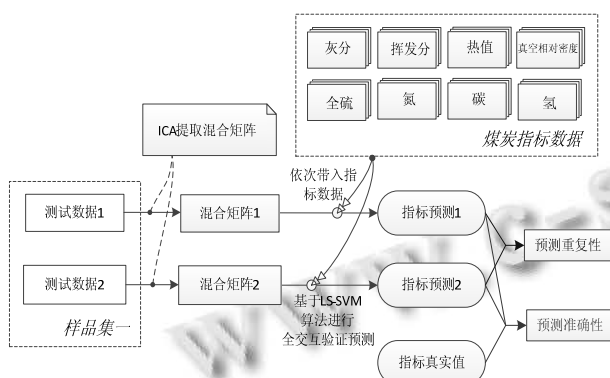


图 7 光谱数据分析流程图

光谱数据 LS-SVM 算法预测的主要结果如图 8-图 15 所示.

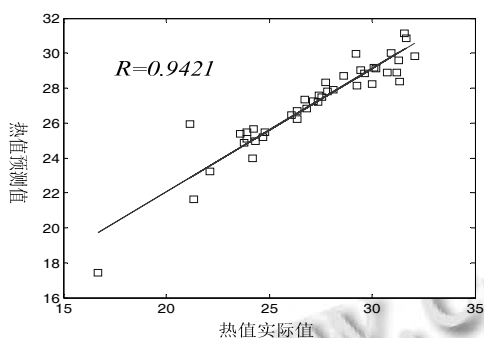


图 8 热值预测结果(R 为相关系数)

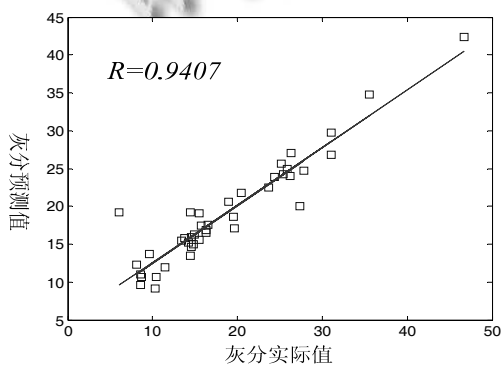


图 9 灰分预测结果

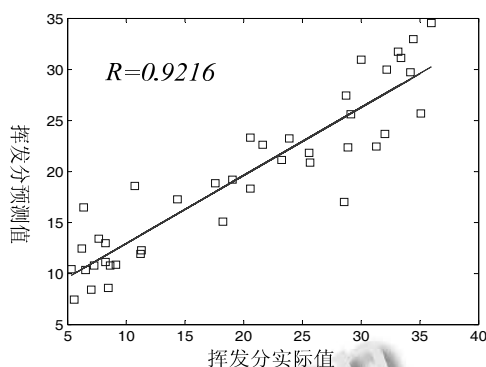


图 10 挥发分预测结果

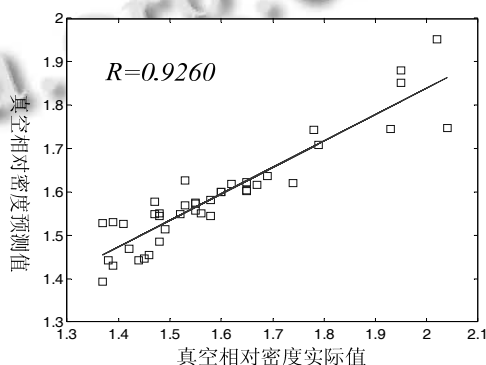


图 11 真空相对密度预测结果

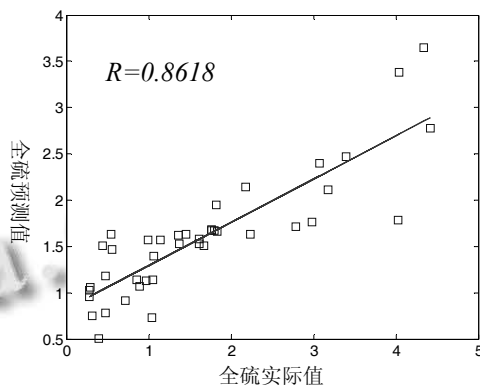


图 12 全硫预测结果

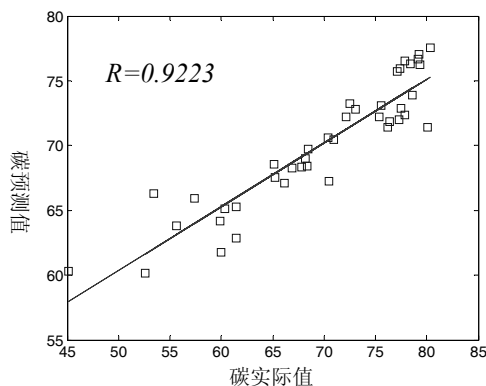


图 13 碳含量预测结果

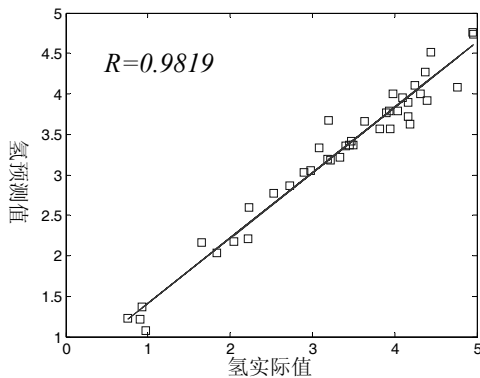


图 14 氢含量预测结果

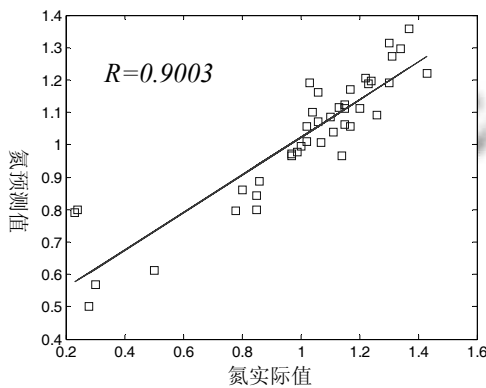


图 15 氮含量预测结果

从表 3 和表 4 可以看出, 使用 ICA+LS-SVM 预测算法, 可以获得了较好的预测效果, 对于热值、灰分、挥发分、真空相对密度、碳含量、氢含量 6 个指标的预测相关系数均达到 0.90 以上; 但是对于全硫含量、氮含量的预测效果相对较差, 预测相关系数在 0.80~0.90 之间. 研究中, 为了便于比较, 同时采用经典的 MLR(多元线性回归)算法对煤炭指标进行了预测, 从表 3~4 可以看出采用 MLR 算法仅能较好地预测挥发分、氢含量、氮含量三个指标, 其余指标的预测相关系数在 0.40~0.85 之间, 预测效果并不理想, 预测结果的重复性不好, 这主要是由于漫反射光谱数据离散性较大, 而 MLR 属于线性模型, 这种数据离散性有可能被 MLR 模型在预测结果中放大, 表现出来就是预测结果的重复性误差较大. 因此, 通过对比可以看出, 相对于经典 MLR 算法而言, LS-SVM 算法有效解决了漫反射光谱数据离散的问题, 提高了煤炭指标预测的准确性和重复性.

表 3 仪器针对各指标预测结果的准确性评价

煤炭指标	预测准确性评价					
	MLR 算法			LS-SVM 算法		
	相关系数	平均误差	最大误差	相关系数	平均误差	最大误差
热值	0.71	2.64	10.22	0.94	0.95	4.79
灰分	0.72	6.51	25.30	0.94	2.06	13.11
挥发分	0.91	3.43	11.03	0.92	3.59	11.55
真空相对密度	0.84	0.08	0.21	0.93	0.06	0.29
全硫含量	0.41	2.99	11.44	0.86	0.53	2.25
碳含量	0.74	4.22	28.77	0.92	3.53	15.15
氢含量	0.94	0.30	1.34	0.98	0.22	0.68
氮含量	0.91	0.08	0.60	0.90	0.09	0.56

表 4 仪器针对各指标预测结果的重复性评价

煤炭指标	预测重复性评价			
	MLR 算法		LS-SVM 算法	
	平均误差	最大误差	平均误差	最大误差
热值	3.54	12.14	0.68	3.75
灰分	8.85	30.34	1.74	7.70
挥发分	4.53	13.89	3.22	11.73
真空相对密度	0.09	0.41	0.06	0.22
全硫含量	3.82	13.82	0.30	1.21
碳含量	9.65	41.30	1.82	9.98
氢含量	0.64	2.17	0.22	0.69
氮含量	0.26	0.79	0.09	0.48

5 结论

研究中, 对 41 个不同质量指标的标准煤样进行了定量分析预测, 考察了在相同粒径的条件下 Hadamard 近红外光谱对煤炭指标的预测能力, 取得了良好预测效果. 总的来看, 使用 Hadamard 近红外光谱技术进行煤炭指标的快速检测是一种可行的方法, 可以满足实际工作中煤质分析的需要, Hadamard 光谱的分辨率问题(分辨率为 8nm, 相比于傅立叶光谱偏低)并未对指标预测产生实质性影响. 基于 Hadamard 变换技术的近红外光谱仪具有更好的稳定性、抗振性和经济性, 更加适合恶劣工况条件下的分析检测, 具备在线分析检测的潜力, 在应用前景上, 较其它煤质无损分析技术具有一定优势.

同时, 基于 Hadamard 变换技术的下一代光谱仪已处于研发中, 新的光谱分辨率可达到 1nm 以下, 已经基本达到傅立叶光谱的分辨率水平, 可以获取更加微

观的光谱细节信息,使得光谱数据的特征性得以进一步提升,有助于得到更好的分析预测结果。另外,从技术的扩展角度而言,HNID-2G煤质分析仪器平台也具有较好的扩展潜力,因为深色物质的测量是漫反射光谱检测的一个难点,如果能够测好深色物质,那么检测浅色物质就显得相对容易。在农产品、食品、烟草等质量监测领域,检测对象多是浅色物质,因此可以预见,将Hadamard近红外检测技术应用在这些领域,应该会有更好的效果。

参考文献

- 1 张孝亮.快速煤质分析仪在集装站煤质管理中的应用.煤质技术,2010,(5):42-44.
- 2 伍秀玲.煤质工业分析指标传统法与全自动工业分析仪方法的比较.煤质技术,2008,(4):35-37.
- 3 李凤瑞,肖宝兰,唐玉国,许思传.应用近红外光谱方法在线测定煤中水分.电站系统工程,2003,19(6):19-20.
- 4 李凤瑞,唐玉国,肖宝兰.应用近红外光谱分析技术测量煤质发热量.电站系统工程,2004,20(3):19-20.
- 5 肖宝兰,李凤瑞.国内外关于燃煤NO_x形成、还原及减排措施的研究进展.发电设备,2004,18(6):350-352.
- 6 李凤瑞,唐玉国,肖宝兰.近红外光谱分析技术预测煤质挥发分含量模型的研究.热能动力工程,2003,18(6):582-584.
- 7 王开明,蔡斌.煤质分析在配煤检测中的作用.煤炭加工与综合利用,2010,(3):28-29.
- 8 王友壮,顾炯,薛伟超.热电厂煤质在线分析的应用研究.煤,2010,19(8):30-32.
- 9 Aisa D, Babucci E, Barocchi F, Cunsolo A. The development of the BRISP spectrometer at the Institut Laue-Langevin. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 2005, 544(3): 620-642.
- 10 Fu DJ, Walker KA, Sung K. The portable atmospheric research interferometric spectrometer for the infrared, PARIS-IR. Journal of Quantitative Spectroscopy and Radiative Transfer, 2007, 103(2): 362-370.
- 11 Wallrabe U, Solf C, Mohr J, Korvink JG. Miniaturized fourier transform spectrometer for the near infrared wavelength regime incorporating an electromagnetic linear actuator. Sensors and Actuators A: Physical, 2005, (123): 459-467.
- 12 Hsu CC, Chen MC, Chen LS. Intelligent ICA-SVM fault detector for non-Gaussian multivariate process monitoring. Expert Systems with Applications, 2010, 37(4): 3264-327.