

文章编号:1671-251X(2015)01-0062-05 DOI:10.13272/j.issn.1671-251x.2015.01.016
李明,李翠,雷萌.一种煤炭近红外光谱数据预处理方法研究[J].工矿自动化,2015,41(1):62-66.

一种煤炭近红外光谱数据预处理方法研究

李明, 李翠, 雷萌

(中国矿业大学 信息与电气工程学院, 江苏 徐州 221008)

摘要:针对煤炭原始近红外光谱数据中存在噪声的问题,提出了基于 De-SNV 与小波阈值去噪组合的煤炭近红外光谱数据预处理方法。采用缺省软阈值法进一步对经过 Savitzky-Golay 平滑和 De-SNV 处理的光谱数据去噪,并分别建立了水分、灰分和挥发分的 PLS 校正模型,通过分析模型的预测性能对该方法的有效性进行评估。实验结果表明,经过该方法预处理的光谱数据所对应的 PLS 校正模型性能明显优于使用原始光谱数据所建立的 PLS 校正模型,水分、灰分和挥发分的 PLS 校正模型的预测均方根误差分别降低至 0.007 07,0.040 8,0.008 66,决定系数分别提高至 0.858 7,0.743 8,0.778 5。

关键词:煤炭光谱;近红外光谱;小波阈值去噪;Savitzky-Golay 平滑;去势-标准正态变换;PLS 校正模型

中图分类号:TD67 文献标志码:A 网络出版时间:2015-01-05 16:37

网络出版地址:<http://www.cnki.net/kcms/detail/32.1627.TP.20150105.1637.016.html>

Research of a data preprocessing method for near infrared spectrum of coal

LI Ming, LI Cui, LEI Meng

(School of Information and Electrical Engineering, China University of Mining and Technology,
Xuzhou 221008, China)

Abstract:For noise existed in original near infrared spectral data of coal, a data preprocessing method for near infrared spectrum of coal was proposed based on De-SNV and wavelet threshold denoising. Spectrum data processed by Savitzky-Golay smoothing and De-SNV was further processed by default soft threshold denoising method. Then PLS calibration models of moisture, ash and volatile were established. Effect of the method was evaluated by analyzing predicting performance of the models. The experiments show that performance of PLS model based on spectrum data processed by the method is much better than the one based on original spectrum data. The root-mean-square errors of prediction of the three PLS calibration models are decreased to 0.007 07, 0.040 8, 0.008 66 respectively, and the determination coefficients are increased to 0.858 7, 0.743 8, 0.778 5.

Key words: coal spectrum; near infrared spectrum; wavelet threshold denoising; Savitzky-Golay smoothing; detrending-standard normal variate; PLS calibration model

0 引言

煤样原始近红外光谱数据除了含有样品的特征信息外,还包含一些干扰信号,如杂散光等^[1],因此需要对数据进行预处理,尽可能去除干扰,优化光谱

信息,提高光谱信号的分析能力,为光谱的特征提取及校正模型的建立奠定基础^[2]。

目前近红外光谱数据预处理方法主要有标准化、中心化、一阶/二阶求导、多元散射校正、傅里叶变换等^[3]。标准化和中心化可增强光谱间的差异,

收稿日期:2014-06-26;修回日期:2014-11-13;责任编辑:李明。

基金项目:高等学校博士学科点专项科研基金资助项目(20110095110011);江苏省自然科学基金资助项目(BK20130207)。

作者简介:李明(1962—),男,安徽淮南人,教授,博士,主要研究方向为智能检测系统,E-mail:liming@cumt.edu.cn。

在一定程度上降低随机噪声,但不具有针对性,也不能消除散射干扰;一阶/二阶求导可有效消除基线漂移,放大光谱谱峰特征信息,同时也放大了随机噪声;多元散射校正主要用于消除光谱中的散射干扰,但不能有效消除光谱中的随机噪声;傅里叶变换只能提供信号的整体频域特征,信号中频率和时间之间缺少一一对应关系^[4],导致预测模型的精确度和稳定性较差。

小波分析方法是一种有效的信号处理方法,在信号去噪领域有着广泛的应用。该方法提供了灵活可变的窗口,通过小波变换将信号按不同的分辨率尺度分解成一系列不同频率的块信号。根据这一特点,可对特殊频率范围的噪声进行滤波处理^[5]。本文提出一种基于 De-SNV (Detrending-Standard Normal Variate, 去势-标准正态变换) 与小波阈值去噪组合的光谱数据预处理方法,在对数据进行 Savitzky-Golay 平滑处理和散射校正的基础上,采用小波阈值方法实现光谱深度去噪。

1 煤炭样本选取

实验借助某国家实验室,来自不同地区的煤炭样本中选取具有代表性的 97 个样本,严格按照 GB 474—2008《煤样的制备方法》要求进行制样,粒度均达到 0.2 mm 级别。在煤炭样本的工业分析中,水分、灰分和挥发分含量的测定均严格遵照参考文献^[6]规定的步骤和要求。采用 Antaris II 傅里叶变换近红外光谱仪采集煤炭样本的近红外光谱,其具体参数:扫描范围为 3 800~10 000 cm⁻¹,波长点数为 1 609,分辨率为 4 cm⁻¹。对每个煤炭样本重复扫描 64 次,取平均光谱作为所测样品的原始光谱。图 1 为所选煤炭样本的近红外光谱,其中 T 为吸光度。

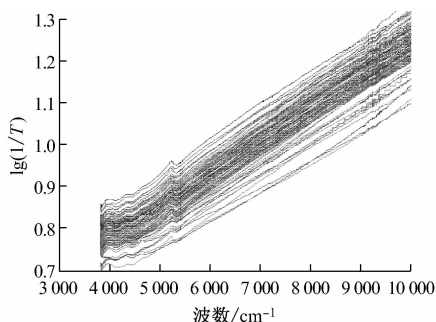


图 1 煤炭样本的近红外光谱

2 实验方法

2.1 De-SNV

SNV 用来校正因散射而引起的光谱差异,其基本思想:在近红外光谱中各个波长点的吸光度应满足一定的分布规律(例如正态分布),在此基础上对任意一条光谱进行校正。计算过程如下:

计算第 i 个样本各波长点光谱吸光度的均值 $\bar{T}_i$ 以及该光谱标准差 S_i :

$$\bar{T}_i = \frac{\sum_{k=1}^m T_{i,k}}{m} \quad (1)$$

$$S_i = \sqrt{\frac{\sum_{k=1}^m (T_{i,k} - \bar{T}_i)^2}{m-1}} \quad (2)$$

对该光谱进行 SNV 校正:

$$T_{i,SNV} = \frac{T_{i,k} - \bar{T}_i}{S_i} \quad (3)$$

式中: $T_{i,k}$ 为第 i 个样本在第 k 个波长点的光谱吸光度; m 为波长点数, $1 \leq m \leq 1\,609$; $T_{i,SNV}$ 为经 SNV 校正后的光谱吸光度。

SNV 常与去势一起使用,简称 De-SNV。De-SNV 可消除非平稳时间序列信号存在的趋势,在一定程度上消除基线漂移。

2.2 小波阈值去噪

2.2.1 小波阈值去噪原理

对噪声信号进行小波变换后得到一系列小波系数,其中有用信号对应的小波系数幅值大,但数量少;噪声信号的小波系数分布一致,幅值小,但数量较多。根据这一原理寻找一个合适的阈值 R 对原信号去噪。主要步骤如下:

(1) 选取母小波,并对含噪信号进行小波分解,得到一系列小波系数。

(2) 选取合适的阈值,并对小波系数进行阈值处理。

(3) 对最后得到的小波系数进行小波逆变换,重构原信号^[7]。

2.2.2 阈值处理方法

设 W 为原小波系数, W' 为处理后的小波系数,常用的阈值处理方法包括硬阈值和软阈值 2 种,分别为

$$W' = \begin{cases} W & |W| \geq R \\ 0 & |W| < R \end{cases} \quad (4)$$

$$W' = \begin{cases} (|W| - R) \operatorname{sgn} W & |W| \geq R \\ 0 & |W| < R \end{cases} \quad (5)$$

软、硬阈值法应用广泛,但都有固有缺点,即在 R 处不连续或有固定偏差,在重构原信号时会产生一定的偏差^[8]。

2.2.3 阈值估计方法

阈值 R 的估计是小波阈值去噪中的关键一步。若 R 取值过大,会将一些有用信号误当作噪声滤除,对建模不利;若 R 取值过小,则不能将噪声彻底滤除。常用的阈值估计方法包括以下几种。

(1) 无偏似然估计阈值法(Rigrsure 规则)。对于风险向量 r' ,其元素为 $r_j = [n - 2j + (n - j)\omega_j + \sum_{k=1}^j \omega_k]/n$, n 为小波系数的个数; $j = 1, 2, \dots, n$ 。取 r' 中最小值 $r_{\min}$ 对应的小波系数为 $\omega_{\min}$,则无偏似然估计阈值为

$$R = \sigma \sqrt{\omega_{\min}} \quad (6)$$

式中: σ 为信号所含噪声标准差。

(2) 最小最大值法(Minimaxi 规则)。该方法选取的阈值是固定阈值的一种,计算公式为

$$R = \begin{cases} \sigma(0.393\ 6 + 0.182\ 9 \frac{\ln n}{\ln 2}) & n \geq 32 \\ 0 & n < 32 \end{cases} \quad (7)$$

(3) Brige-Massart 阈值法。该方法指定一个分解层数 c ,对于 $c + 1$ 及更高层,所有小波系数保留;对于第 $l(1 \leq l \leq c)$ 层,保留绝对值最大的 n_l 个系数。

$$n_l = M(l + 2 - c)^a \quad (8)$$

式中: M, a 为经验系数。

(4) 缺省阈值法。阈值计算公式为

$$R = F \sqrt{2 \lg p} \quad (9)$$

式中: F 为根据原信号估计得到的噪声强度; p 为信号长度。

在实际应用中,对于如何选择阈值策略尚没有统一的准则,只有根据具体问题来选择合适的阈值进行去噪处理^[9]。

3 实验结果及分析

煤炭样本依次采用 Savitzky-Golay 平滑、De-SNV、小波阈值去噪的组合方法进行预处理。经 Savitzky-Golay 平滑和 De-SNV 处理后的光谱如图 2 所示。

由于 SNV 是对每条光谱分别进行校正,所以经过 De-SNV 处理后,光谱特征信息更加明显,波

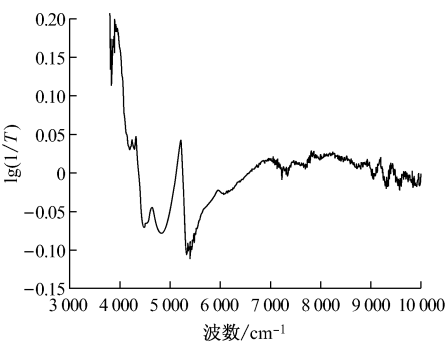


图 2 经 Savitzky-Golay 平滑和 De-SNV 处理后的光谱峰、波谷等信息也更加清晰,在抑制散射效应以及线性平移的同时,加强了光谱之间的差异性,使得光谱特征信息更加丰富,更有利于建立模型。

利用近红外光谱原始数据以及经过 De-SNV 处理后的数据,分别建立水分、灰分以及挥发分的 PLS(Partial Least Squares, 偏最小二乘法)校正模型,并用决定系数和预测均方根误差(RMSEP)对模型进行评价^[10],结果见表 1。

表 1 近红外光谱数据的 PLS 校正模型评价结果

模型	水分		灰分		挥发分	
	决定系数	RMSEP	决定系数	RMSEP	决定系数	RMSEP
A	0.836 0	0.008 30	0.713 5	0.043 9	0.713 4	0.012 49
B	0.851 3	0.008 12	0.732 7	0.044 6	0.748 9	0.010 63

注: A 代表原始数据的 PLS 模型, B 代表经 De-SNV 处理后数据的 PLS 模型。

进行小波阈值去噪时,选用 db4 小波。本实验选取最小最大值法,并分别在软、硬阈值条件下与缺省阈值法和 Brige-Massart 阈值法的去噪效果进行比较。选择信噪比(SNR)和 RMSEP 并结合去噪后的波形作为评价指标^[11]。通常认为,SNR 越高, RMSEP 越小,则表示去噪效果越好。

图 3 为 1 号样本的原始光谱数据及采用 Brige-

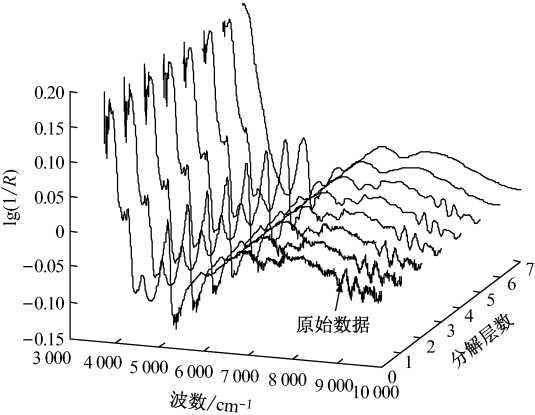


图 3 不同分解层的 Brige-Massart 去噪结果

Massart 软阈值法进行 1—7 层小波去噪处理后的结果,表 2 为不同分解尺度下对应的 SNR 和 RMSEP。通过综合比较,当分解层数为 3 时,波形在高频部分较清晰,低频部分较平滑,SNR 为 30.990,RMSEP 为 0.002 2,去噪效果最好^[12]。

表 2 Brige—Massart 软阈值去噪结果

指标	1 层	2 层	3 层	4 层	5 层	6 层
SNR	43.022	34.787	30.990	28.886	26.213	20.800
RMSEP	0.000 7	0.001 5	0.002 2	0.002 7	0.003 5	0.006 1

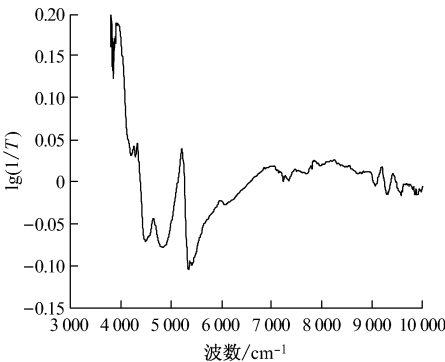
用同样的方法分别确定缺省阈值法和最小最大值法在软、硬阈值条件下的最优分解层数。去噪结果如图 4 所示,SNR 和 RMSEP 见表 3。可看出当选用缺省软阈值法时,SNR 最大,RMSEP 最小,去

噪效果最好。

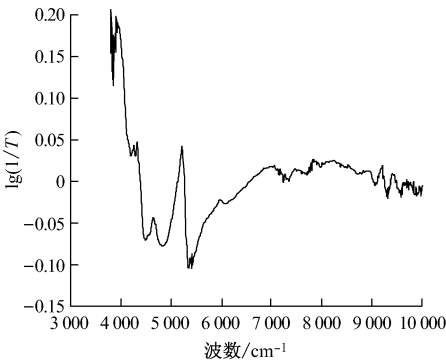
表 3 不同小波阈值方法的 SNR 和 RMSEP

阈值估计 方法	SNR		RMSEP	
	软阈值	硬阈值	软阈值	硬阈值
缺省阈值法	36.382	34.946	0.001 1	0.001 5
Brige—Massart 阈值法	30.990	33.066	0.002 2	0.001 8
最小最大值法	31.633	27.990	0.001 7	0.003 0

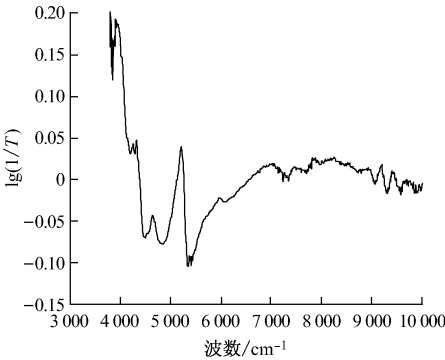
用组合方法对光谱数据进行预处理,并分别建立水分、灰分和挥发分的 PLS 校正模型,得到的模型评价结果见表 4,可看出 PLS 校正模型性能均有所提高。



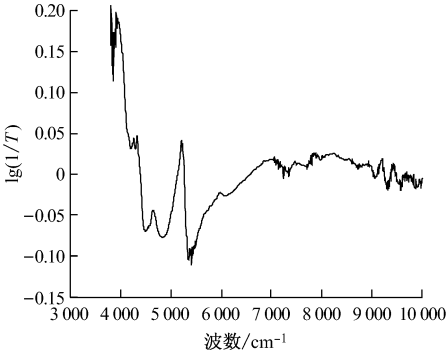
(a) Brige—Massart 软阈值法



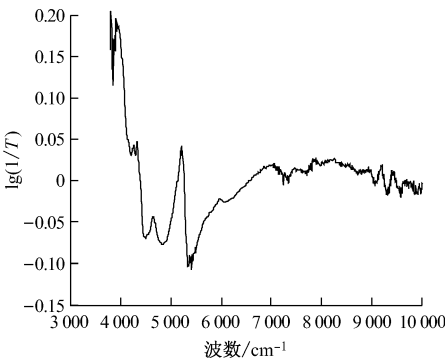
(b) Brige—Massart 硬阈值法



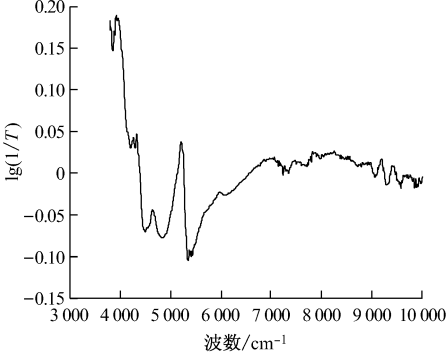
(c) 缺省软阈值法



(d) 缺省硬阈值法



(e) 最小最大软阈值法



(f) 最小最大硬阈值法

图 4 不同小波阈值方法的去噪结果

表 4 经组合方法处理后数据的 PLS 校正模型评价结果

水分		灰分		挥发分	
决定系数	RMSEP	决定系数	RMSEP	决定系数	RMSEP
0.858 7	0.007 07	0.743 8	0.040 8	0.778 5	0.008 66

4 结论

(1) 经过 De-SNV 处理后,煤炭近红外光谱特征信息更加明显,但同时高频段的噪声也更加明显,需要进行深度处理。

(2) 在 Savitzky-Golay 平滑和 De-SNV 校正的基础上,对光谱数据进行缺省软阈值法去噪后,水分、灰分和挥发分 PLS 校正模型的决定系数分别提高到 0.858 7,0.743 8 和 0.778 5,模型的精度显著提高。

(3) 采用原始以及经组合方法预处理后的光谱数据分别建立 PLS 校正模型。通过对比预测性能可知,经过组合方法预处理后,水分 PLS 校正模型的决定系数和预测均方根误差变化不大,灰分与挥发分 PLS 校正模型性能提升明显。

参考文献:

[1] 陈鹏强,陆辉山,闫宏伟. 基于近红外光谱的煤粉样品定量检测研究[J]. 工矿自动化,2013,39(8):68-71.

[2] LEWIS A T,JONES K,LEWIS K E,*et al.* Detection of Lewis antigen structural change by FTIR spectroscopy [J]. Carbohydrate Polymers, 2013, 92(2):1294-1301.

[3] KIM D W,LEE J M,KIM J S. Application of near infrared diffuse reflectance spectroscopy for on-line measurement of coal properties[J]. Korean Journal of Chemistry Engineering,2009,26(2):489-495.

[4] PALLUA J D,PEZZEI C,ZELGER B,*et al.* Fourier

transform infrared imaging analysis in discrimination studies of squamous cell carcinoma [J]. Analyst, 2012,137(17):3965-3974.

[5] HUANG Z R,SHA S,RONG I Q,*et al.* Feasibility study of near infrared spectroscopy with variable selection for non-destructive determination of quality parameters in shell-intact cottonseed [J]. Industrial Crops and Products,2013,43:654-660.

[6] GB/T 212—2008 煤的工业分析方法[S].

[7] NASRI M, NEZAMABADI-POUR H. Image denoising in the wavelet domain using a new adaptive thresholding function[J]. Neurocomputing, 2009, 72 (4/5/6):1012-1025.

[8] HEDAOO P, GODBOLE S. Wavelet thresholding approach for image denoising [J]. International Journal of Network Security & Its Applications, 2011,3(4):16-21.

[9] TESCH S,RENTROP K H,OTTO M. Coal Analysis by application of the partial least squares method to infrared spectra[J]. Fresenius Journal of Analytical Chemistry,1992,344:206-208.

[10] 李华,王菊香,郭恒光,等. 光谱预处理方法对混胶近红外定量模型影响的研究[J]. 分析科学学报,2010, 26(5):551-554.

[11] SMITH B M, GEMPERLINE P J. Wavelength selection and optimization of pattern recognition methods using the genetic algorithm[J]. Analytical Chemistry Acta,2000,423:167-177.

[12] DONG W K,JONG M L,JAE S K. Application of near infrared diffuse reflectance spectroscopy for on-line measurement of coal properties [J]. Korean Journal of Chemistry Engineering, 2009, 26 (2) : 489-495.