

基于多重多元回归的焦炭质量预测模型

张进春, 吴超

中南大学资源与安全工程学院, 长沙 410083

摘要 焦炭质量预测是焦化企业进行焦炭质量控制的重要方法。在诸多生产因素已固定的条件下, 焦炭质量主要取决于原料煤性质。用煤质指标预测焦炭质量是焦炭质量预测的重要方法。考虑配合煤煤质指标和焦炭质量指标的多元性和相关性, 采用多重多元回归分析技术, 将焦炭质量的各指标作为一个整体, 建立配合煤煤质指标对焦炭质量的预测模型。基于偏最小二乘回归思想, 采用预测误差平方和 (PRESS) 和预测的方差验证回归模型的预测能力。根据实际焦炭生产建立的焦炭质量预测多重多元回归模型的显著性检验表明, 该模型具有较高的预测能力。应用多重多元回归技术建立的焦炭质量预测模型对指导焦化企业的焦化生产, 优化配煤和加强焦炭质量控制具有重要的现实意义。

关键词 焦炭质量预测; 多重多元回归; 预测误差平方和; 预测方差

中图分类号 TD984

文献标识码 A

文章编号 1000-7857(2010)12-0079-06

Coke Quality Prediction Model Based on Multivariate Regression

ZHANG Jinchun, WU Chao

School of Recourse and Safety Engineering, Central South University, Changsha 410083, China

Abstract The prediction of coke quality in a coke quality controlling process is important for coke production facilities. With many production conditions such as equipments, parameters, fixed, the coke quality mainly depends on the quality of coal. Using the quality index of coal to forecast the quality of coke is an important method of coke quality prediction. Because the relations between quality indexes of the blended coal and the coke take a multiple nature and are mutually related, so all quality indexes of the coke are taken as a whole and then the multivariate regression is used to establish the coke quality prediction model. In order to assure the model a good performance in predicting, firstly, the fitting accuracy of the regression model is tested, then, the Prediction Error Square Summation (PRESS) and Prediction Variance are used to test the model based on the partial least-squares algorithm. According to a practical coke production, a coke quality prediction model based on multivariate regression is built. The significant test shows that the model is correct and the predicting test shows that this model is accurate in prediction. The practical application of the method shows that the coke quality prediction model based on multivariate regression is reliable for coke production in blending coal optimally and in coke quality control.

Keywords coke quality prediction; multivariate regression; prediction error square summation; prediction variance

0 引言

随着高炉的大型化和富氧喷吹煤粉等技术的开发应用, 高炉生产对焦炭质量的要求愈来愈高。决定高炉焦炭质量的因素很多, 但在现行箱式焦炉的生产中, 诸多因素已固定, 焦炭质量主要取决于原料煤的性质^[1]。因此通过掌握炼焦煤性质, 用煤质指标预测焦炭质量, 使焦炭质量稳定在目标范围

内, 已成为焦炭质量控制与管理的重要方法。

国内外研究者提出了不同的焦炭质量预测方法^[2-11], 概括起来, 这些方法大体可以分为两类: 一类是基于回归分析, 建立配合煤煤质指标和焦炭质量指标的回归模型进行预测; 另一类是运用神经网络等智能算法, 以配合煤的煤质指标作为输入, 焦炭的质量指标作为输出, 建立煤质指标和焦炭质量

收稿日期: 2010-01-19

基金项目: 国家自然科学基金项目 (50974132)

作者简介: 张进春, 博士研究生, 研究方向为质量工程与安全工程, 电子信箱: zjc@hpu.edu.cn; 吴超 (通信作者, 中国科协所属全国学会个人会员登记号: E520001585M), 教授, 研究方向为安全与环保, 电子信箱: wuchao@mail.csu.edu.cn

指标间的预测模型进行预测。然而上述焦炭质量预测模型存在着一些问题。一方面,用配合煤煤质指标建立的焦炭质量预测模型应是一个多元输入多元响应问题,且多响应指标间具有相关性,而多数焦炭质量预测的多元回归模型对此却未予考虑,仅针对各焦炭质量指标分别建立多元线性回归模型,这样生硬割裂焦炭质量指标之间的关系不符合生产实际。另一方面,由于神经网络训练程度很难把握,许多算法往往过分强调训练结果而牺牲泛化结果,致使网络的过拟合问题严重^[12],即所建立的焦炭质量预测模型看似解释了模型运算中配合煤煤质指标和焦炭质量指标之间的关系,但不能根据新的配合煤煤质指标对焦炭质量指标进行有效预测。原因在于这些智能算法将随机波动也纳入到模型拟合中,导致所建立的模型泛化能力并不高,同时,限于现场工作人员的技能,这些复杂算法在具体生产实践中难以有效应用,不具备较高的应用性。

为了避免上述问题,本文采用多重多元回归分析技术,建立焦炭质量预测的多重多元回归模型。相对多元回归分析模型建立的多元输入一元响应回归模型,多重多元回归分析模型建立的是多元输入多元响应回归模型,且基于多元响应的相关性进行回归拟合^[13]。多重多元回归模型的参数估计与检验基于最小二乘法原理,考虑控制随机波动,从而使回归模型具有较好的拟合优度。因此,应用多重多元回归分析技术,建立配合煤煤质指标与焦炭质量指标间的多重多元回归分析模型是合适的。为了进一步检验回归模型是否具有较好的泛化能力,本文基于偏最小二乘思想,采用基于预测误差平方和与预测的方差进行回归模型的预测能力检验,从而保证建立的焦炭质量预测模型具有较高的有效性。

1 多重多元回归分析的基本原理

1.1 多重多元回归分析模型及参数估计

设有 m 个自变量和 p 个因变量的 n 组观测值,分别用数据矩阵 X, Y 表示:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix} \quad Y = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1p} \\ y_{21} & y_{22} & \cdots & y_{2p} \\ \vdots & \vdots & & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{np} \end{bmatrix}$$

记 β 和 E 分别为回归系数与残差,表示为

$$\beta = \begin{bmatrix} \beta_{01} & \beta_{02} & \cdots & \beta_{0p} \\ \beta_{11} & \beta_{12} & \cdots & \beta_{1p} \\ \vdots & \vdots & & \vdots \\ \beta_{m1} & \beta_{m2} & \cdots & \beta_{mp} \end{bmatrix} = \begin{bmatrix} \beta'_{(0)} \\ \beta'_{(1)} \\ \vdots \\ \beta'_{(m)} \end{bmatrix} \quad E = \begin{bmatrix} \varepsilon_{11} & \varepsilon_{12} & \cdots & \varepsilon_{1p} \\ \varepsilon_{21} & \varepsilon_{22} & \cdots & \varepsilon_{2p} \\ \vdots & \vdots & & \vdots \\ \varepsilon_{n1} & \varepsilon_{n2} & \cdots & \varepsilon_{np} \end{bmatrix} = \begin{bmatrix} \varepsilon'_{(1)} \\ \varepsilon'_{(2)} \\ \vdots \\ \varepsilon'_{(n)} \end{bmatrix}$$

则多重多元线性回归模型可表示为

$$\begin{cases} Y = (I_n^T X) \beta + E = C \beta + E \\ \varepsilon_{(i)} \sim N_p(0, \Sigma) \quad i=1, 2, \cdots, n; \text{相互独立} \end{cases} \quad (1)$$

其中, Σ 为残差 $\varepsilon_{(i)}$ 的方差-协方差矩阵, C 为 $n \times (m+1)$ 设计矩阵。

根据最小二乘法,计算回归参数 β 的最小二乘估计为

$$\hat{\beta} = (C^T C)^{-1} C^T Y \quad (2)$$

1.2 多重多元回归模型的检验

1.2.1 回归模型的检验

回归模型的检验是回归分析的重要工作。对于回归模型,通常首先进行直观的残差分析,初步判断回归模型是否正确,然后进行显著性检验。其中直观检验主要包括残差的正态概率图检验和残差与预测变量之间关系的检验,如果回归模型是正确的,则各响应分量残差则应该服从或渐进服从正态分布,且各响应分量的残差与响应无关,二者之间不应呈现某种特殊形态^[14]。

对多重多元回归模型进行显著性检验,需要考查任一自变量 $X_{(i)}$ 对因变量 Y 的各个分量 $Y_{(j)}$ 影响是否显著的问题就是对于回归系数 $\beta_i = (\beta_{(1)}, \beta_{(2)}, \cdots, \beta_{(i)}, \cdots, \beta_{(m)})$ 进行假设检验:

$$H_0: \beta_{(1)} = \beta_{(2)} = \cdots = \beta_{(m)} = 0 \quad (3)$$

相应的检验统计量为

$$\frac{MSS/f_{MSS}}{ESS/f_{ESS}} \sim F(f_{MSS}, f_{ESS}) \quad (4)$$

其中, MSS 为回归平方和, ESS 为剩余平方和, f_{MSS}, f_{ESS} 为相应自由度。

1.2.2 模型预测能力的检验

回归模型的作用主要在于预测和控制,因此建立的回归模型必须保证具备较高的预测性能,这意味着所建立的回归模型既不能失拟,也不能过拟,即必须保证一定的拟合优度。为此,采用预测误差平方和 (PRESS) 和预测的方差 R_p^2 验证回归模型的预测能力。

PRESS 是基于残差尺度度量的一种有效的模型预测检验方法。为计算 PRESS,先选择一个观测 i ,然后用其余 $n-1$ 个观测拟合回归模型,并用这个回归方程预测保留地观测 y_i ,记这个预测值为 $\hat{y}_{(i)}$,则可得该点预测误差为 $e_{(i)} = y_i - \hat{y}_{(i)}$ 。对每一观测值重复此过程,则可得预测误差平方和为

$$PRESS = \sum_{i=1}^n e_{(i)}^2 \quad (5)$$

PRESS 统计量不需要按其定义进行计算,而采用^[14]

$$PRESS = \sum_{i=1}^n \left(\frac{e_i}{1 - h_{ii}} \right)^2 \quad (6)$$

式中, h_{ii} 为帽子矩阵 $H = C(C^T C)^{-1} C^T$ 的对角元素。预测的 R_p^2 为

$$R_p^2 = 1 - \frac{PRESS}{SST} \quad (7)$$

式中, SST 为总偏差平方和。

2 基于多重多元回归焦炭质量预测

2.1 模型的建立

焦炭指标主要包括焦炭的灰分 $A_d(\%)$ 、硫分 $S(\%)$ 、抗碎强度 $M_{40}(\%)$ 和耐磨强度 $M_{10}(\%)$ 等。这些指标直接影响高炉冶炼的冶炼效果,是评价焦炭质量的重要指标。配合煤的 A_d 、 S 、挥发分 $V_{daf}(\%)$ 和黏结性指数 G 直接影响焦炭性能。配合煤中的 A_d 和 S 几乎全部转入焦炭之中,因此,配合煤的 A_d 、 S 与

表 3 2009 年度各月份焦炭质量预测
Table 3 Coke quality prediction in each month in 2009

时间	配合煤煤质指标				焦炭质量指标预测值				焦炭质量指标实测值			
	$A_d / \%$	$V_{daf} / \%$	$SI / \%$	G	$A_d / \%$	$SI / \%$	$M_{40} / \%$	$M_{10} / \%$	$A_d / \%$	$SI / \%$	$M_{40} / \%$	$M_{10} / \%$
2009-01	9.74	29.03	0.64	77.11	12.3195	0.5352	69.4571	7.6793	12.65	0.53	72.89	7.51
2009-02	9.79	29.33	0.61	80.56	12.4897	0.5363	74.0972	9.0216	12.17	0.61	73.95	9.08
2009-03	9.98	28.19	0.52	72.31	13.6893	0.5626	83.4095	8.3465	13.65	0.56	84.47	8.45
2009-04	9.84	29.32	0.54	63.79	12.6636	0.5057	67.6137	8.0459	12.61	0.53	68.46	8.04
2009-05	10.02	28.45	0.65	73.13	13.5774	0.5908	78.9478	8.2298	13.52	0.60	77.34	8.04
2009-06	9.98	28.44	0.58	74.72	13.5372	0.5727	81.6126	8.5736	13.95	0.59	75.56	8.64
2009-07	9.82	28.37	0.65	79.27	12.8577	0.5682	75.7260	7.4041	12.83	0.57	76.28	7.43
2009-08	9.97	28.12	0.62	65.55	13.4731	0.5721	74.2079	6.8055	13.66	0.58	72.90	6.69
2009-09	9.90	28.65	0.58	73.15	13.1378	0.5524	76.7716	8.1788	13.46	0.56	76.20	8.32
2009-10	9.81	28.10	0.61	70.08	74.755	-0.335	7.2540	6.236	13.28	0.53	71.20	6.30

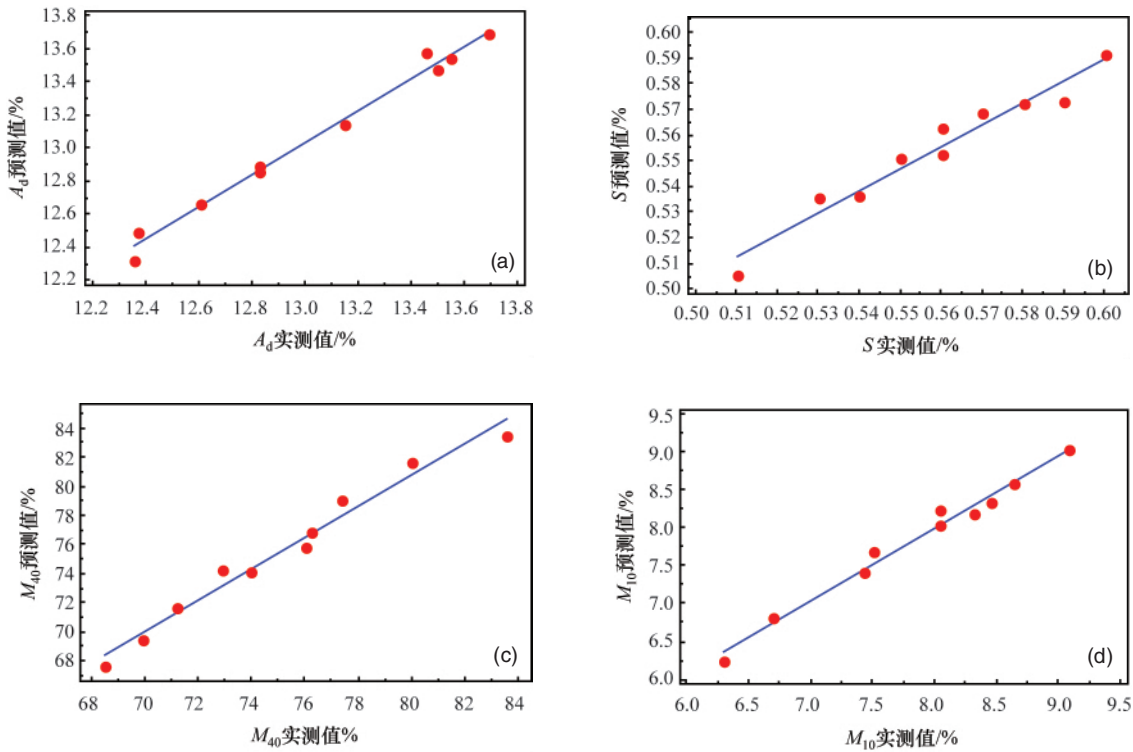


图 3 2009 年度各月焦炭质量预测值与实测值
Fig. 3 Coke quality prediction value and actually value in each month in 2009

3 结论

焦炭质量指标是一个多元质量特性问题,各个指标间具有相关性,采用多重多元回归分析技术,将焦炭质量指标作为一个整体,建立配合煤煤质指标对焦炭质量的预测模型。模型检验表明,建立的焦炭质量多重多元回归预测模型是合适的;模型预测能力检验也说明所建立的多重多元回归模型具有较高的预测能力。但需要指出的是,焦炭质量的多重多元回归预测模型是基于煤源、煤质相对稳定和其他焦化条件不变的条件下进行的,各个企业应根据企业的实际情况建立

焦炭质量的多重多元回归预测模型;同时要考核特定焦化条件下焦炭质量指标与配合煤煤质指标的线性关系,某些情况下应对配合煤煤质指标进行数据变换,再建立多重多元回归预测模型。总之,焦炭质量预测的多重多元回归预测模型对指导焦化企业进行优化配煤和焦炭质量控制具有现实意义。

参考文献 (References)

[1] 周师庸. 炼焦煤性质和高炉焦炭质量 [M]. 北京: 冶金工业出版社, 2005.

- Zhou Shiyong. Properties of coking coal quality of coke for the blast furnace[M]. Beijing: Metallurgical Industry Press, 2005.
- [2] McKenzie F, Hockings K, Livingstone P. Modeling of coal blending to predict coke quality [C]//Proc AIE 8th Australian Coal Science Conf. Sydney, 1998: 101-106.
- [3] Díez M A, Alvarez R, Barriocanal C. Coal for metallurgical coke production: Predictions of coke quality and future requirements for cokemaking[J]. *International Journal of Coal Geology*, 2002, 50(1-4): 389-412.
- [4] Zhang Q, Wu X C, Feng A Z, *et al.* Prediction of coke quality at Baosteel [J]. *Fuel Processing Technology*, 2004, 86(1): 1-11.
- [5] Alvarez R, Díez M A, Barriocanal C, *et al.* An approach to blast furnace coke quality prediction[J]. *Fuel*, 2007, 86(14): 2159-2166.
- [6] 杨秀华, 丁玉献. 运用回归分析对炼焦生产的预测与控制 [J]. 煤质技术, 2004(6): 25-26.
Yang Xiuhua, Ding Yuxian. *Coal Quality and Technology*, 2004 (6): 25-26.
- [7] 王光辉, 范程, 田文中, 等. 焦炭质量预测方法的研究 [J]. 武汉科技大学学报, 2007, 30(1): 37-40.
Wang Guanghui, Fan Cheng, Tian Wenzhong, *et al.* *Journal of Wuhan University of Science and Technology*, 2007, 30(1): 37-40.
- [8] 高志芳, 单晓云, 朱书全, 等. 二元线性回归法预测焦炭强度的研究[J]. 选煤技术, 2006, 6: 1-3.
Gao Zhifang, Shan Xiaoyun, Zhu Shuquan, *et al.* *Coal Preparation Technology*, 2006, 6: 1-3.
- [9] 刘俊, 张学东, 刘宏, 等. 基于 BP 神经网络的焦炭质量预测[J]. 燃料与

- 化工, 2006, 37(6): 12-16.
- Liu Jun, Zhang Xuedong, Liu Hong, *et al.* *Fuel & Chemical Process*, 2006, 37(6): 12-16.
- [10] 郭一楠, 巩敦卫, 程健. 基于分布式神经网络的焦炭质量预测模型[J]. 中国矿业大学学报, 2005, 34(4): 514-517.
Guo Yinan, Gong Dunwei, Cheng Jian. *Journal of China University of Mining & Technology*, 2005, 34(4): 514-517.
- [11] 崔庆安, 何桢, 崔富新. 基于支持向量机的焦炭质量预测模型 [J]. 化工自动化及仪表, 2006, 33(1): 28-31.
Cui Qing'an, He Zhen, Cui Fuxin. *Control and Instrument in Chemical Industry*, 2006, 33(1): 28-31.
- [12] 覃光华, 李祚泳. BP 网络过拟合问题研究及应用[J]. 武汉大学学报, 2006, 39(6): 55-58.
Qin Guanghua, Li Zuoyong. *Engineering Journal of Wuhan University*, 2006, 39(6): 55-58.
- [13] Johnson R A, Wichern D W. 实用多元统计分析 [M]. 北京: 清华大学出版社, 2001.
Johnson R A, Wichern D W. *Applied multivariate statistical analysis* [M]. Beijing: Tsinghua Press, 2001.
- [14] Montgomery D C. 实验设计与分析[M]. 北京: 人民邮电出版社, 2007.
Montgomery D C. *Design and analysis of experiment*[M]. Beijing: Post & Telecom Press, 2007.
- [15] 姚昭章, 郑明东. 炼焦学[M]. 北京: 冶金工业出版社, 2005.
Yao Zhaozhang, Zheng Mingdong. *Coke making* [M]. Beijing: Metallurgical Industry Press, 2005.

(责任编辑 代丽)

第一届中国植物修复生物学学术研讨会

时间: 2010年8月13-16日
地点: 福建·厦门

主办单位: 中国植物生理学会

电 话: 0592-6190997 电子信箱: ygzhu@rcees.ac.cn
网 址: <http://www.iue.cas.cn/xwzx/xshd/201004/P020100420587076763028.pdf>