

偏最小二乘法在炭黑工艺数据解析中的应用

黎金明 李梦龙* 印家健 姜 微

(四川大学化学学院 成都 610064)

摘 要 针对目前炭黑行业生产主要以经验为主的不利境况, 为了改进炭黑生产技术, 采用了偏最小二乘法对工艺数据进行处理, 建立了工艺参数与产品指标之间的数学模型, 解析了工艺参数对炭黑吸碘值和 DBP 吸油值的影响程度, 得出工艺生产的主控参数和辅助控制参数。利用计算结果详细分析了如何根据生产数据信息调整工艺, 提高产品质量。应用该算法结果对实际生产进行指导, 获得了比较满意的成效, 提高了生产优等炭黑的能力。

关键词 炭黑 偏最小二乘法 吸碘值 DBP 吸油值 数据解析

Application of Partial Least Squares in Parameter Analysis for Carbon-Black Production

Li Jinming, Li Menglong*, Yin Jiajian, Jiang Wei

(Faculty of Chemistry, Sichuan University, Chengdu 610064)

Abstract In terms of the production of carbon black by means of experience principally, and for promoting carbon black production technology, partial least squares(PLS) is applied to the analysis of the data from production process, and the mathematics model between parameters and targets is set up. After the affected extent of each parameter to iodine specific surface area(ISSA) and DBP absorption is parsed, primary parameters and assistant parameters are found out. Based on the computed results, how to adjust production parameters for improving product quality of carbon-black is discussed in detail. Subsequently, the results from PLS are applied to guide the actual production, satisfactory effectiveness is gained, and the ability for producing high quality carbon black is improved.

Key words Partial least squares, Carbon-black, Iodine specific surface area, DBP absorption, Parameter Analysis

炭黑是由许多烃类物质(固态、液态或气态)经不完全燃烧或裂解生成的。炭黑是橡胶制品(主要是轮胎)的重要补强剂和填充剂^[1]。橡胶工业的技术更新和不断突破, 相应对炭黑的质量要求更高, 从而刺激炭黑生产技术从经验式生产转向自动化、数字化生产, 同时达到精细化(如纳米炭黑、专用特种炭黑等)生产^[1,2]。在烃类物质质量稳定的情况下, 炭黑质量的好坏主要受生产工艺参数的影响。炭黑生产工艺参数约 50 个, 都分别归属于温度、压力、流量三种类型中, 可见工艺数据为高维数据; 炭黑质量衡量标准中的全指标有 13 个, 故又为多指标问题。目前, 年生产能力上万吨级的炭黑生产工艺, 大都实现了自动化控制, 比如 DSC 控制系统, 能够在计算机

黎金明 男, 29 岁, 硕士生, 现从事化学计量学研究。*联系人, E-mail: liml@scu.edu.cn

国家教育部骨干教师基金资助项目

2003-12-30 收稿, 2004-03-15 接受

上调节各工艺参数^[2,3], 但如何根据炭黑产品信息调节各参数方面的技术处理并不清晰, 甚至有些工厂还是主要凭经验; 在实际生产中也只建立了简单的线性回归模型, 能做粗略的工况预测, 却缺乏先进的数据处理方法、工艺优化建模方法来改进炭黑工艺。这对于提高炭黑质量, 以及满足橡胶工业及其它相关产业的发展极为不利。鉴于此, 本文采用偏最小二乘法(partial least squares, 简称 PLS), 可以去掉原始变量的相关性, 过滤原始数据的噪声, 抽取特征变量, 压缩高维数据进行降维, 将参数和目标的特征反映在 PLS 载荷投影图中。从建立的数学模型可以分析出对目标影响比较大的参数, 通过工艺数据解析结果改进炭黑的生产技术, 提高生产优质炭黑的能力。

1 方法原理

1.1 基本思路

设有 m 个变量对 s 个目标有影响, m 个变量分别为 x_1, x_2, \dots, x_m , s 个目标分别 y_1, y_2, \dots, y_s 。原始数据先按标准化进行预处理, 经过处理后的变量达到权重相同, 均值为 0, 方差为 1。采用欧氏距离法(Euclidean distance)删除离群样本, 通过 Fisher 比率法进行特征变量筛选。经过样本筛选变量筛选后的数据共有 n 组, 即有 n 个样本, 每个样本代表一种生产工况, 则原始数据集 $X=[x_{ij}](i=1, 2, \dots, n; j=1, 2, \dots, m); Y=[y_{ij}](i=1, 2, \dots, n; j=1, 2, \dots, s)$, X 为各自变量(工艺参数)构成的数据矩阵, Y 为各目标(炭黑吸碘值和 DBP 吸油值)构成的数据矩阵, PLS 的实质在于将 Y 中不利于建模的噪声排除掉, 将有用的信息部分通过 Y 的负载与得分两个矢量和 X 的负载和得分相互交叉回归, 以求得兼顾 X 和 Y 都拟合得比较好的负载 PLS 成分^[4]。PLS 的迭代算法过程可参阅参考文献^[3,6], 现给出 X 和 Y 的几种关系如下:

(1) X 和 Y 的外部关系为

$$X_{n \times m} = T_{n \times m} P_{m \times m}^T + E \quad Y_{n \times s} = U_{n \times s} Q_{s \times s}^T + F \quad (1)$$

式中 T 、 U 为得分矩阵, P 、 Q 为载荷矩阵, E 、 F 为系统模型不能解释的随机误差矩阵。

(2) X 和 Y 的内部关系为:

$$u_h = v_h t_h + e \quad (2)$$

式中 h 为主成分数, e 为回归的残差, 此式说明要求 t_h 与 u_h 为最大重迭或相关性为最大。

(3) X 和 Y 的混合关系为:

$$Y = TVQ^T + F \quad \min \|F\| \quad (3)$$

在迭代算法中, X 和 Y 得到相互都是最好的得分, 从而获得较好的内部关系。

1.2 PLS 成分数目的确定

采用模型交叉有效性检验(cross validation)法确定 PLS 成分数目。首先应用留 K 法(leave-K-out)计算预报残差平方和(prediction residual error sum of squares, PRESS), 就是在 n 个原始数据样本中留下 k 个样本, 用其余的训练样本建模, 然后利用该模型预报被留下的 k 个样本, 计算它们的残差平方和。轮流留下 k 个样本, 依次建模、预报、计算残差平方和, 最后总加起来就是 PRESS^[4-8]。本文中取 K 为 1, 即留一法。

$$\text{PRESS} = \sum_{i=1}^n \sum_{j=1}^s (y_{ij} - y_{ij}^p)^2 \quad (4)$$

式中 y^p 为被留下的样本被预报的值, y 为该样本的原始数值。 n 为样本数, s 为目标的个数。

PLS 成分从少到多, 每一种成分数目, 都进行 n 次留一法计算该种数目成分的 PRESS, 比较 PRESS 减少的趋势, 当 PRESS 减小到达到某一阈值时, 以此为判据确定 PLS 成分。根据模型分析各变量对各目标的影响, 找到参数与目标之间的数学关系, 从而定量调节工艺参数, 使工艺参数向生产优质产品方向变化。本文采用 Matlab 语言编写了计算程序。

2 数据及预处理

数据来源于中橡集团炭黑工业研究设计院炭黑生产实验车间原始记录, 共 112 个样本, 分为优等品类(S1), 一等品类(S2)和合格品类(S3)。根据炭黑生产经验, 参与优化控制的操作变量共有 7 个, 即天然气流量、原料油流量、一次急冷水流量、二次急冷水流量、添加剂量、干燥机碳黑出口温度和造粒机功率, 分别以 x_1 、 x_2 、 x_3 、 x_4 、 x_5 、 x_6 、 x_7 表示。以炭黑吸碘值和吸油值为产品质量衡量指标, 分别以 y_1 和 y_2 表示。

原始数据先按标准化^[3]进行预处理, 经过处理后的变量达到权重相同, 均值为 0, 方差为 1。采用欧氏距离法(Euclidean distance)和 Fisher 权重法^[3]对原始数据进行处理。利用 Fisher 权重法进行变量筛选。经计算, 各变量 Fisher 权重如表 1 所示。

表 1 各变量的 Fisher 权重
Tab.1 Fisher weight of each variable

变量	x_1	x_2	x_3	x_4	x_5	x_6	x_7
Fisher 权重 ^a	0.0039	0.2253	0.0011	0.0007	0.0641	0.1116	0.0459
Fisher 权重 ^b	0.0157	0.0535	0.0194	0.0127	0.0684	0.0074	0.0091
Fisher 权重 ^c	0.0280	0.0506	0.0270	0.0078	0.0009	0.0879	0.0207

a: 对 S1、S2 的 Fisher 权重; b: 对 S1、S3 的 Fisher 权重; c: 对 S2、S3 的 Fisher 权重

从表 1 可以看出, 7 个变量的 Fisher 权重时大时小, 说明其中每个变量对样本分类影响都不能忽略。如果按 Fisher 权重从大到小进行排序, x_2 总排在前 2 位, 其它变量次序变化都比较大。因此在以下模式识别中 $x_1 \sim x_7$ 入选为特征变量。经过样本筛选变量筛选后, 得到含 7 个特征变量的样本 86 个。

3 结果与讨论

利用偏最小二乘法对数据进行处理, 借助交叉有效性检验法, 确定 PLS 成分数为 2。经过偏最小二乘法中的迭代算法, 得到目标矩阵 Y 的载荷因子矩阵 Q^T 和工艺参数矩阵 X 的载荷矩阵 p^T :

$$Q^T = \begin{bmatrix} 0.4697 & 0.8828 \\ -0.9319 & 0.3628 \end{bmatrix}$$

$$p^T = \begin{bmatrix} 0.1067 & -0.5050 & 0.0620 & 0.3161 & 0.8404 & -0.1054 & 0.2394 \\ 0.3375 & 0.1399 & 0.8969 & 0.3560 & 0.0206 & 0.1727 & 0.2399 \end{bmatrix}$$

再由工艺参数矩阵的得分矢量 t_1 和 t_2 就可以根据式(5)、(6)计算出：

$$t_1 = 0.1067x_1 - 0.5050x_2 + 0.0620x_3 + 0.3161x_4 + 0.8404x_5 - 0.1054x_6 + 0.2394x_7 \quad (5)$$

$$t_2 = 0.3375x_1 + 0.1399x_2 + 0.8969x_3 + 0.3560x_4 + 0.0206x_5 + 0.1727x_6 + 0.2399x_7 \quad (6)$$

同样，目标矩阵 Y 的得分矢量 u_1 和 u_2 为：

$$u_1 = 0.4697y_1 + 0.8828y_2 \quad u_2 = -0.9319y_1 + 0.3628y_2 \quad (7)$$

两个得分矢量间的内部相关矩阵为：

$$V = \begin{bmatrix} 0.6244 & 0 \\ 0 & 0.7141 \end{bmatrix}$$

从而，两个得分矢量之间的内部关系式表示为：

$$u_1 = 0.6424t_1 + e \quad u_2 = 0.7141t_2 + e \quad (8)$$

从 Y 矩阵的载荷矩阵 Q^T ，可以计算出目标矢量与得分矢量的关系：

$$y_1 = 0.4697u_1 - 0.9319u_2 \quad y_2 = 0.8828u_1 + 0.3628u_2 \quad (9)$$

从式(9)中得出，如果要增加 y_1 (吸碘值)，工艺条件的调整应该是 u_1 朝正方向变化， u_2 朝负方向变化；如果增加 y_2 (吸油值)，则 u_1 、 u_2 都朝正方向变化。为同时满足增大 y_1 和 y_2 ，可以得出 u_1 应增加；对于 u_2 ，由于 u_2 减小的幅度(幅度系数为-0.9319)比其增加幅度(幅度系数为 0.3628)大，所以 u_2 总趋势是减少。式(8)表明， u_1 与 t_1 、 u_2 与 t_2 在 PLS 法迭代计算中是内部线性相关的，所以可以通过调节 t_1 、 t_2 来满足 u_1 、 u_2 需要的条件。从式(5)看到， t_1 主要与 x_2 负相关，与 x_5 正相关，即减少 x_2 、增加 x_5 就能达到增加 y_1 (吸碘值)的目的；同时有 t_2 主要与 x_1 、 x_3 、 x_4 正相关，特别是与 x_3 的相关系数达 0.8969，减小 x_1 、 x_3 、 x_4 也能直接增加 y_1 (吸碘值)。同样，调整 x_1 、 x_2 、 x_3 、 x_4 、 x_5 也能有效得到 y_2 (吸油值)。从以上数据关系解析结果得出， x_1 、 x_2 、 x_3 、 x_4 、 x_5 是影响炭黑产品质量的关键因素，尤其 x_2 、 x_3 、 x_5 应是主要调控参数， x_6 、 x_7 只作为辅助控制参数。从以上数学关系，就可解决如何根据产品中间分析信息，即 y_1 (吸碘值)、 y_2 (吸油值)来调节工艺参数 x_1 、 x_2 、 x_3 、 x_4 、 x_5 、 x_6 、 x_7 ，使炭黑吸碘值和吸油值被控制在优质品指标范围内。

另外，PLS 方法还可通过载荷矩阵投影图来分析参数与目标的关系。由于载荷投影图反映了变量之间的关系。因此，只需将目标载荷矢量 q_1 对 q_2 以及参数载荷矢量 p_1 对 p_2 同时作图(图 1)。

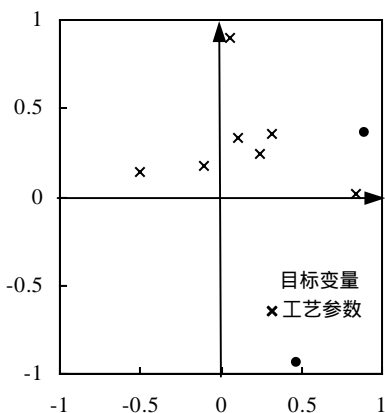


图 1 PLS 荷载投影图

Fig.1 The projection diagram of PLS loadings

1— y_1 ; 2— y_2 ; 3— x_1 ; 4— x_2 ; 5— x_3 ; 6— x_4 ; 7— x_5 ; 8— x_6 ; 9— x_7

从图 1 中得出：1 和 2 的变化主要由 $7(x_5)$ 决定，这主要反映在 PLS_1 轴上。 $4(x_2)$ 、 $7(x_5)$ 控制了 PLS_1 的变化， $5(x_3)$ 主要控制了 PLS_2 的变化， $3(x_1)$ 、 $6(x_4)$ 对 PLS_2 作用也不小。因此，3~7 是对产品影响比较大的参数，其中 4、5、7 是关键参数，这与直接分析数学模型得出的结论一致。由于 1、2 处于不同方向，如果要提高 y_2 减小 y_1 ，就可直接增加 3、5、6 来实现；增加 7 可以同时增加 y_1 和 y_2 ；减小 4，可以同时减小 y_1 和 y_2 。可以看出，PLS 载荷投影图的分析比数学模型分析更容易更直观，但是达不到对操作变量的定量控制。而建立的数学模型更能表达参数与目标的清晰关系，更能把握由产品检测数据 y_1 和 y_2 来决定工艺参数($x_1 \sim x_7$)的调节幅度和变化方向。只有将二者结合，才能快速准确诊断工艺情况，科学地调整工艺参数，达到炭黑生产的优化控制，提高优等炭黑产品的比率。

为了验证以上数据解析的可靠性，预先设定 5 个目标样本，按吸碘值、吸油值的大小分为 3 类，根据 PLS 数学模型计算出对应 5 个样本的工艺参数，见表 2。然后利用表 2 中工艺参数调整工艺进行实验，并对所生产的炭黑进行吸碘值、吸油值测试，结果列于表 3 中。

表 2 5 个目标样本对应的工艺参数
Tab.2 The technics parameters corresponding to five target samples

编号	类别	天然气流量 /(Nm ³ /h)	原料油流量 /(kg/h)	一次及冷水 流量/(kg/h)	二次急冷水 流量/(kg/h)	添加剂量 /(kg/h)	干燥机碳黑出 口温度/	造粒机功率 /kW
1	1	506	2870	3458	742	0.006	258	24.2
2	1	505	2870	3480	742	0.005	244	25.3
3	2	509	2800	3449	631	0.008	221	25.5
4	3	511	2830	3471	665	0.065	257	26.60
5	3	506	2840	3498	630	0.065	251	26.24

表 3 实验值与预定值的比较
Tab.3 The comparison between experiment value and enactment value

编号	类别	吸碘值/(g/kg)			吸油值/(10 ² m ³ /kg)		
		实验值	预定值	相对误差/%	实验值	预定值	相对误差/%
1	1	121	120	0.83	122	125	2.40
2	1	117	120	2.50	119	122	2.46
3	2	126	126	0.00	123	121	1.65
4	3	114	117	2.56	115	116	0.86
5	3	121	118	2.54	117	117	0.00

由表 3 可知，吸碘值、吸油值实验值与预定值比较接近，相对误差在 2.56% 以内，达到实际生产中的要求。因此，PLS 方法比较好地解决了炭黑生产模型构建问题，利用模型对工况进行预测，克服盲目生产造成损失。

4 结论

为了改变炭黑长期经验式生产现状，提高炭黑生产技术水平，本文采用了偏最小二乘法对炭黑工艺数据进行处理，得出天然气流量、原料油流量、一次急冷水流量、二次急冷水流量以及添加剂量是影响炭黑产品质量的关键因素，尤其原料油流量、一次急冷水流量和添加剂量应是主要调控参数，干燥机碳黑出口温度和造粒机功率只作为辅助控制参数。同时，通过生产实验验证了工艺参数与目标之间的 PLS 数学模型，结果表明，实验值与预定值相对误差在 2.56% 以内，达到实际生产中的要求，提高了优等炭黑产品的比率。

参考文献

- [1] 李炳炎. 炭黑生产与应用手册. 北京: 化学工业出版社, 2000: 122~212.
- [2] M Janik. Fuel and Energy, 2001, 42(1): 20~24.
- [3] 俞汝勤. 化学计量学导论. 长沙: 湖南教育出版社, 1991: 345~346.
- [4] A Lazraq, R Cle'roux. J. Chemometrics., 2001, 15: 523~536.
- [5] A Phatak, S D Jong. J. Chemometrics, 1997, 11: 311~338.
- [6] X Wang, U Kruger, Lennox. Control Engineering Practice, 2003, 11(6): 613~632.
- [7] B Li, J Mprris, B Martin. Chemom. Intell. Lab. Syst., 2002, 64(1): 79~89.
- [8] M C DENHAM. J. Chemometrics. 1997, 11: 39~52.