

有序 Gram-Schmidt 正交化在研究 无取向硅钢质量中的应用^①

朱尔一 邓志威 黄本立
(化 学 系)

摘要 无取向硅钢工艺是一个多变量系统,影响产品质量的素多.本工作从生产现场收集了 147 组数据,应用有序 Gram-Schmidt 正交化等多元分析方法处理数据,找出影响质量的主要因素,根据 PRESS 判据排除多噪音的因素,建立预报能力较强的模型,并从分析中找出提高质量的措施.

关键词 无取向硅钢, Gram-Schmidt 正交化, PRESS 判据

无取向硅钢是一种电机用钢.其基本生产流程为:首先由炼钢,连铸和热轧得到符合成分要求的热轧钢板,然后经过酸洗,冷轧,再通过连续退火炉.在退火过程中需将碳脱到较低值,以保证低铁损和低磁时效.最后还要在表面加以处理,使带钢具有防腐性和好的冲片性以及用作电动机芯必要的电器绝缘性.

无取向硅钢的生产过程是一个复杂的多元系统,目前尚无法从理论上建立准确反应内在规律的机理模型.该钢种的主要质量指标是钢的铁损 P_{15} 和成品的含碳量.而成品牌号高低的判定主要是根据铁损 P_{15} 值的大小,铁损 P_{15} 值越低,该电工钢的牌号就越高.影响该电工钢牌号(质量)的因素很多,主要是钢的化学成分,轧机状况,均热退火炉状况,气氛的露点等工艺条件.在实际生产中,由于种种原因,有时有原材料相同(同一炉次的钢)的情况下,生产出钢的成品牌号时高时低,不很稳定.针对这种情况,我们应用有序 Gram-Schmidt 方法,总结实际生产中数据,以找出生产该钢种成品牌号波动的原因或生产过程中主要工艺参数对成品牌号的影响,并找出提高该钢种成品牌号生产过程优化工艺条件.

我们从现场收集数据共 147 组,每一组数据包含 16 个因素和目标变量 P_{15} (铁损),这 16 个因素是,钢中化学元素的含量有 C、Si、Mn、P、S、Al、Cu,以及后工序工艺参数有 ZR(轧机状况), SF_{1-6} , SF_7 , SF_{8-11} , SF_{12} , SF_{13} (段均热退火炉炉温), PP (露点), $H_2\%$ (氢气含量), T (水温)等.在作建模和进一步分析前,我们对原始数据进行了预处理,使各变量数据的均值为 0,均方差为 1.

1 分析方法

目前在多元数据分析中广泛采用 PLS^[1]和 PCR^[2]两种有序正交化方法,本工作采用了另一种有序正交化方法,有序 Gram-Schmidt 方法.在有序 Gram-Schmidt 方法中原始数据矩阵 X 被分解为正交矩阵 T ,而 T 矩阵中的隐变量一方面为两两正交,另一方面又按一定顺序排

① 本文 1992-11-21 收到

列,与目标变量相关性最大的变量排在前面,排在后面的隐变量被认为包含较多成分的噪声,因此可通过删除一些含噪声多的隐变量,使噪声与信息能够很容易地分离.此外,有序 Gram-Schmidt 方法所得到的正交矩阵 T 中的隐变量,可分别对应于各原始变量,因此在分离噪声的同时,可除去影响不显著的变量,以达到筛选变量的目的.有序 Gram-Schmidt 正交化方法在做变量筛选计算过程中,用 PRESS 判据来确定变量的剔除和引入,不需事先作任何假设,因而使用较方便. PRESS (Prediction Sum of Squares)^[3] 的定义式为

$$\text{PRESS} = \sum_i (Y_i - \hat{Y}_{i,-i})^2 \quad (1)$$

其中 $\hat{Y}_{i,-i}$ 为第 i 个样本不参加建模时,模型的预测值,因此 PRESS 与普通的残差平方和不同,普通的残差平方和表示模型拟合的好坏,而 PRESS 表示模型的预报能力, PRESS 值越小表示模型的预报能力越强.因此用 PRESS 值作为判据可确定模型变量的取舍.

用有序 Gram-Schmidt 正交化方法得到一组正交化矢量 t_1, t_2, \dots, t_n 和以下特点,首先矢量 t_1, t_2, \dots, t_n 是两两正交的,因此矢量(或变量)之间没有信息重迭.其次,矢量 t_1, t_2, \dots, t_n 按与目标 y 的相关性大小排列,前面的矢量等与目标的相关性大,包含有用信息多,噪声的成分少,而后面的矢量与目标 y 的相关性小,含有较多的噪声成分.因此这样就便于噪声与有用信息的分离.因为这样我们很容易地使用 PRESS 判据,当将矢量 t_1, t_2, \dots, t_n 按顺序逐个引入模型,并分别求出对应的 PRESS 值,根据 PRESS 值达到最低值确定引入变量个数 h ,并将前面矢量 t_1, t_2, \dots, t_h 引入模型,后面的矢量 t_{h+1}, \dots, t_n 为包含噪声成分较多的量而不引入模型.由于前面的变量 t_1, t_2, \dots, t_h 没有包括 x_{h+1}, \dots, x_n 的信息,所以 x_{h+1}, \dots, x_n 也可以被删除,从而起到变量筛选的作用.

有序 Gram-Schmidt 正交化方法算式如下:

第一步

1) 将 X 中的矢量按与 y 的相关系数的大小排序,相关系数大的矢量排在前头,得到一组矢量 $x_1^{(1)}, x_2^{(1)}, \dots, x_n^{(1)}$.

2) 取 $P_{11} = 1, t_1 = x_1^{(1)}$

3) $P_{1i} = (t_1^T x_i^{(1)}) / (t_1^T t_1) \quad i = 2, 3, \dots, n$

4) $x_i^{(2)} = x_i^{(1)} - P_{1i} t_1$

第二步(循环)

对于 $k = 2, 3, \dots, n$

1) 将矢量按与目标 y 的相关系数的大小排序,得到一组矢量 $x_k^{(k)}, x_{k+1}^{(k)}, \dots, x_n^{(k)}$.

2) $P_{kk} = 1, t_k = x_k^{(k)}$

3) $P_{ki} = (t_k^T x_i^{(k)}) / (t_k^T t_k) \quad i = k+1, \dots, n$

4) $x_i^{(k+1)} = x_i^{(k)} - P_{ki} t_k \quad i = k+1, \dots, n$

经过以上循环计算,可完成矩阵 X 的有序正交化分解.

2 结果与讨论

首先我们用有序 Gram-Schmidt 方法对原始数据矩阵 X 进行正交化分解,得到的正交矢量再逐个引入模型,并计算 PRESS 值,结果见图 1.

从图 1 可知,当变量数为 7 时 PRESS 达到最低值,因此取 7 个变量建模,这 7 个因素对应

于原始变量的平均值和波动方差见表 1.

表 1 影响 P_{15} 的因素
Tab. 1 Factors influencing P_{15}

| | 钢的化学成分(Z) | | | | | ZR | SF_{13} | P_{15} |
|------|-----------|-------|---------|-------|---------|-------|-----------|----------|
| | C | Mn | p | Al | Cu | 轧机 | 炉温℃ | (w/kg) |
| 平均值 | 0.004 7 | 0.25 | 0.007 4 | 0.321 | 0.048 1 | 1.4 | 828.3 | 4.898 |
| 波动方差 | 0.001 6 | 0.011 | 0.002 8 | 0.038 | 0.008 6 | 0.49 | 6.87 | 0.469 |
| 符号 | X_1 | X_2 | X_3 | X_4 | X_5 | X_6 | X_7 | Y |

我们根据 Gram-Schmidt 方法得出变量排列顺序可知,影响 P 主要因素的排序为 1, X_5 ; Cu; 2, X_6 ; ZR; 3, X_2 ; Mn; 4, X_1 ; C; 5, X_7 ; SF_{13} ; 6, X_3 ; P; 7, X_4 ; Al.

有序 Gram-Schmidt 方法求得的数学预测模型为

$$Y = -2.1075 + 54.489X_1 - 10.603X_2 + 17.59X_3 + 1.0951X_4 - 20.329X_5 + 0.0022X_6 + 0.0114X_7$$

从以上得出结果的分析看,影响无取向硅钢铁损值 P_{15} 的主要因素的是钢中 Cu 的含量,而且适当地增加 Cu 的含量有利于降低铁损值 P_{15} . 这个结果从一般的认识来看是有争议的,一般认为硅钢中 Cu 含量 > 0.7% 是有害的,但在这里 Cu 的含量是 0.05% 左右,远小于 0.7%,在这样小的含量范围,Cu 对硅钢

铁损值的影响目前研究并不是很多,日本有人指出在取向硅钢增加一些 Cu 是有利的,但对于无取向硅钢 Cu 的影响还尚未见报道,其次,ZR(轧机状态)对 P_{15} 也有较大影响. 分析结果表明,钢卷过 1 号轧机比过 2 号轧机好,这两个轧机的参数是不同的,另外,适当地增加一点 Mn,降低一点 C、P、Al 有利于降低 P_{15} . 另外第 13 段均热退火炉的温度 SF_{13} 应稍低一些较好.

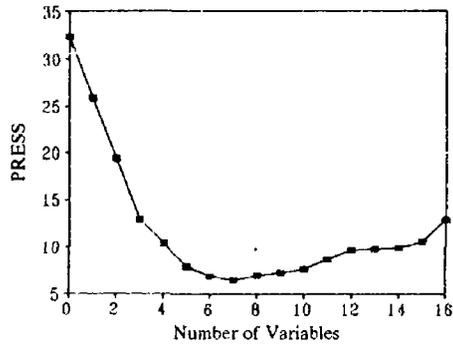


图 1 变量引入数与 PRESS 值关系
Fig. 1 Relation between Number of variables and PRESS

参 考 文 献

- 1 Hoskuldsson A. PLS regression methods. *J. Chemometrics*, 1988, 2: 211~228
- 2 Haaland D M, Thomas E V. PLS method for spectral Analysis. *Anal. Chem.*, 1988, 60: 193~1 202
- 3 Myers R H. *Classical and Modern Regression with Application*. Boston, Massachusetts: Duxbury Press, 1986. 100~111

Apply Order Gram-Schmidt Orthogonalization to Research Quality of Disoriented silicon steel Sheet

Zhu Eryi Deng Zhiwei Huang Benli

(Dept. of Chem.)

Abstract The production of disoriented silicon steel sheet is a multivariate system, in which there are many factors effecting the quality of the products. In this work the data sets included 147 samples are collected from the real production process and the Order Gram-Schmidt orthogonalization method is used to deal with the data in order to find the main factors effecting the quality of products. According to PRESS criteria, the variables with more noise are discarded and the model with high predictive ability is established. The measure for improving the quality of products is determined in the analysis.

Key words Disoriented silicon steel sheet, Gram-Schmidt orthogonalization, PRESS criteria

(上接 364 页)

Screening of Yeast Strains for Inulinase Production and Fermentation Conditions

Zheng Zhonghui Liu Yueying Cai Wenzheng Wei Wenling Zheng Zhicheng

(Dept. of Biol.)

Abstract A strain of *Kluyveromyces* sp. Y-85, a high-yielding strain for producing inulinase was screened from 101 yeast strains. The inulinase production was induced by inulin or xylose. The suitable medium for the enzyme production was composed of 7% Jerusalem artichoke extracted juice, 2% urea and 3% corn steep liquor. The optimal cultivation temperature and initial pH for the enzyme production were 30 °C and 5-6 respectively. Under these conditions, a maximum inulinase activity of 1 403 u/ml was obtained after 24 hours of growth in shake cultures.

Key words Inulinase, *Kluyveromyces* sp., High fructose syrup