

一种新的动态过程数据校正方法*

赵 捧 蒋慰孙

(华东理工大学自动化研究所·上海, 200237)

摘要: 数据校正技术能有效地提高测量数据的质量. 本文提出了一种基于过程模型和统计分析的动态过程数据校正方法, 该方法能同时进行过失误差检测并最终得到各个变量的校正值. 对这种方法进行了仿真研究, 仿真结果证实了该方法的有效性.

关键词: 数据校正; 过失误差检测; 统计分析

A New Method of Dynamic Data Rectification

Zhao Peng and Jiang Weisun

(Research Institute of Automatic Control, East China University of Science and Technology·Shanghai, 200237, P.R. China)

Abstract: Data rectification technology can effectively improve the quality of measurements. In this paper, a new method of dynamic data rectification based on process model and statistical analysis has been proposed. This method can detect gross errors and obtain rectified values of the measurements. Simulation results on CSTR process verified its effectiveness.

Key words: data rectification; gross error detection; statistical analysis

1 引言 (Introduction)

数据校正问题从 60 年代开始, 得到了广泛、深入的研究, 在实际生产过程中也得到了一些应用. 然而现有的数据校正方法绝大部分是面向稳态过程而言的. 实际上一个生产过程或许永远不可能真正处于稳态, 因此采用动态过程模型能更好地近似一个实际过程, 而研究动态过程数据校正方法具有更普遍的意义.

由于动态过程的复杂性, 到目前为止, 动态过程数据校正方法还远未得到充分的发展, 文献报道中的大部分方法都只能处理带有随机误差的情况, 不能同时进行过失误差的检测. 本文探索了一种基于过程模型和统计分析的动态过程数据校正方法, 并对这种方法进行了仿真验证, 结果证实了该方法的有效性.

2 数据校正方法 (Method for data rectification)

本文提出的数据校正方法基本原理如下: 在获取过程输入变量的测量数据后, 依据过程模型, 得到输出变量的计算值, 将这些计算值和实际测量值进行比较, 在无过失误差存在的情况下, 其差值序列 $\eta(j)$ 应近似服从均值为 0 的正态分布.

对于数据窗长度为 n 的随机序列:

$$\{\eta(j) \mid j = k - n + 1, k - n + 2, \dots, k\}, \quad (1)$$

作假设:

H_0 —序列(1)的均值为 0, 即系统正常;

H_1 —序列(1)的均值非 0, 即系统在 k 时刻已存在过失误差.

同时构造统计量

$$U = \frac{\bar{\eta} - 0}{S^* / \sqrt{n}}, \quad (2)$$

其中

$$\bar{\eta} = \frac{1}{n} \sum_{j=k-n+1}^k \eta(j), \quad (3)$$

$$S^* = \sqrt{\frac{1}{n-1} \sum_{j=k-n+1}^k (\eta(j) - \bar{\eta})^2}. \quad (4)$$

由数理统计知识^[1]可知, U 服从自由度为 $n-1$ 的 t 分布. 取假设检验的显著性水平 $\alpha = 0.05$, 即选择参数 $t_{\alpha/2}$, 使

$$P\{|u_i| \geq t_{\alpha/2}(n-1) \mid H_0\} = \alpha (= 0.05). \quad (5)$$

其中, u_i 是 U 的第 i 个分量. 查 t 分布表, 得到临界值 $t_{\alpha/2}$, 从而确定接受区域 $[-t_{\alpha/2}, t_{\alpha/2}]$ 以及否定域 $(-\infty, -t_{\alpha/2})$ 和 $(t_{\alpha/2}, \infty)$. 因此过失误差的检测可根据如下的决策规则进行: 1) 若 U 的各个分量 u_i 均满足 $u_i \leq t_{\alpha/2}$, 则接受假设 H_0 , 即认为系统变量中无

* 国家自然科学基金资助项目(29576232).
本文于 1997 年 7 月 18 日收到, 1998 年 5 月 25 日收到修改稿.

过失误差存在.2) 否则,接受假设 H_1 ,即认为系统变量中存在过失误差,若有几个 u_i 同时满足 $u_i > t_{\alpha/2}$,则可推断是输入变量中带有过失误差;若仅有一个 u_i 满足 $u_i > t_{\alpha/2}$,则可推断是与该 u_i 对应的输出变量带有过失误差.

当确定了带有过失误差的测量变量后,使可考虑采用 Kalman 滤波估计出过失误差的大小.由于 Kalman 滤波是基于模型的技术,因此必须给出过程测量变量的偏差模型.在一般情况下,测量偏差的动态特性是未知的,模拟测量偏差最为简单的方法是把它看作一个常数和噪声之和.基于此,对于系统

$$\begin{cases} X(k+1) = f(X(k), k) + v(k), \\ Z(k+1) = h(X(k+1), k+1) + e(k+1). \end{cases} \quad (6)$$

若已检测到第 n 个测量变量带有过失误差,则可假定

$$\begin{cases} b(k+1) = b(k) + v_b(k), \\ z_n(k+1) = h_n(X(k+1), k+1) + b(k+1) + e_n(k+1). \end{cases} \quad (7)$$

其中 b 表示过失误差.在此基础上,重新构造系统模型为:

$$\begin{cases} \begin{bmatrix} X(k+1) \\ b(k+1) \end{bmatrix} = \begin{bmatrix} f(X(k), k) \\ b(k) \end{bmatrix} + \begin{bmatrix} v(k) \\ v_b(k) \end{bmatrix}, \\ Z(k+1) = h(X(k+1), k+1) + \begin{bmatrix} 0 \\ b(k+1) \\ 0 \end{bmatrix} + e(k+1). \end{cases} \quad (8)$$

依据这一新的系统模型,采用 Kalman 滤波^[1](或扩展 Kalman 滤波)进行滤波计算,估计出过失误差 b

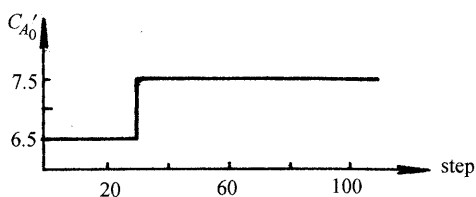


图1 进料浓度示意图
Fig. 1 Feed concentration

的大小,同时得到最终的校正结果.

3 仿真结果及其说明 (Simulation results and illustrations)

以连续搅拌釜反应器(CSTR)中进行的一级不可逆放热反应为例,进行了仿真研究.对文献[2]中给出的动态过程模型进行无因次化处理,得到:

$$\frac{dC_A'}{dt} = 0.01(C_{A_0}' - C_A') - 7.86 \times 10^{12} \cdot \exp(-140.9/T') C_A', \quad (9)$$

$$\frac{dT'}{dt} = 0.01(T_0' - T') + 2.12 \times 10^{12} \cdot \exp(-140.9/T') C_A' - 0.005(T' - 3.4). \quad (10)$$

其中 C_{A_0}' 为无因次化进料浓度; T_0' 为无因次化进料温度; C_A' 为无因次化釜内浓度; T' 为无因次化釜内温度.

依据模型(9),(10),用本文提出的方法作了仿真研究.由于采用长度为 n 的数据窗,在检测到某一变量的统计量超限之时,则一般而言该变量已在一段时间之前带上了过失误差,在本例中把这段时间取为 $\frac{1}{2}n$ 个步长.一旦检测到某一变量的统计量超限,则推断其在前 $\frac{1}{2}n$ 个步长开始带有过失误差.下面给出部分仿真结果.

1) 系统正常运行的情况.

在这种情况下,假定各测量变量均不带有过失误差.系统正常运行时的测量数据由仿真产生.首先给定进料温度和浓度,如图1,2所示;然后利用式(9),(10)计算得到釜内温度和浓度;最后再给输入、输出变量叠加上均值为0、标准差为0.05的正态分

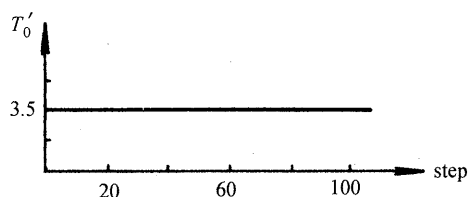


图2 进料温度示意图
Fig. 2 Feed temperature

布随机干扰信号,以此作为各变量的正常测量数据.

依据过程输入变量的测量数据和过程模型计算出过程输出,将输出变量的计算值和实测值相比较,得到差值序列 $\eta(j)$,取数据窗长度 $n = 10$,分别计算统计量 $|u_{C_A}'|$ 和 $|u_T|$.取显著性水平 $\alpha = 0.02$,

查 t 分布表^[3] 得 $|t_{\alpha/2}| = 2.821$.计算结果表明, $|u_{C_A}'|$ 和 $|u_T|$ 均未超限,因此可推断此时输入输出变量中均不带有过失误差.

2) 给 T_0' 加上一个斜坡形式的过失误差.

在 $1 \sim 49$ 步时, T_0' 取其正常测量值;在 $50 \sim 80$

步中,在 T_0' 的正常测量值上叠加一个斜坡信号,即: $T_0' = \text{正常测量值} + 0.1(i - 50)$,其中 i 表示第 i 个时间步长;从第 81 步开始, T_0' 又恢复为其正常测量值。

依据仿真数据分别计算统计量 $|u_{C_A}|$ 和 $|u_T|$ 。结果表明,两个统计量同时从第 61 步开始超限,由于从同一时刻开始同时出现两个过失误差

的概率很小,因此可推断是输入变量从第 56 步开始带有过失误差。依前述基本原则,重新构造系统模型,应用扩展 Kalman 滤波进行滤波计算,得到过失误差的估计值,如图 3 所示。

观察图 3 可以发现,本例中过失误差的估计值和它的真实值之间有一定差值,这可能是由于非线性模型线性化时带来的误差、过失误差的假定模型

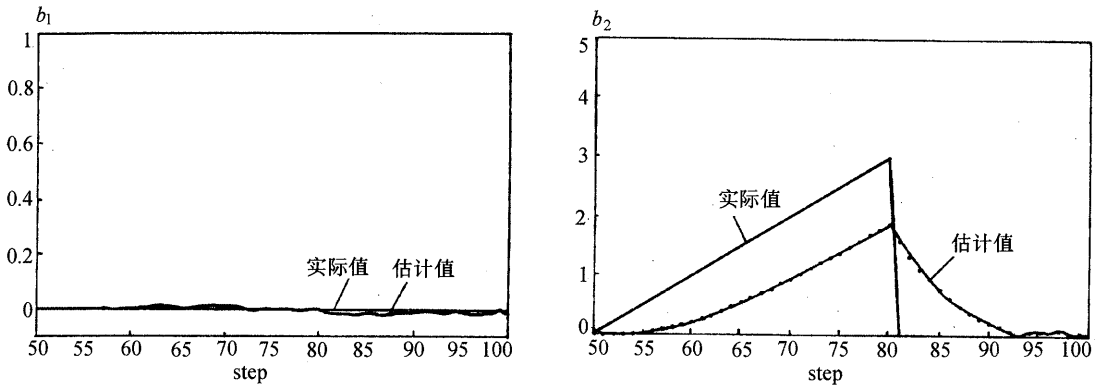


图 3 输入变量存在过失误差时的 b 估计值曲线

Fig. 3 The estimates of b with a gross error existing in the input variable

和实际模型之间的不匹配等原因所致。

4 结语 (Conclusion)

本文提出的数据校正方法直接依据过程模型和输入输出变量的测量值判断过失误差的存在与否,在判断出过失误差后,借助 Kalman 滤波(或扩展 Kalman 滤波)得到过失误差的估计值及各过程变量的校正值。过程变量中无论是否存在过失误差,都可用这种方法实现数据校正。对于无法对过程变量模型作出合理假设的情况,仍可采用该方法进行过失误差检测,一旦检测到过失误差,可提示操作人员去检修传感器,进而排除故障。

参考文献 (References)

- 1 Kalman R E. A new approach to linear filtering and prediction problems. Trans. AME, J. Basic Eng., 1960, 82(3): 34-45
- 2 Liebman M J, Edgar T F and Lasdon L S. Efficient data reconciliation and estimation for dynamic process using nonlinear programming techniques. Comput. Chem. Eng., 1992, 16(10/11): 963-986
- 3 李俊德. 应用数理统计方法. 郑州: 河南科学技术出版社, 1985

本文作者简介

赵 捧 女, 1965 年生, 分别于 1989 年和 1997 年在华东理工大学工业自动化专业获得硕士和博士学位, 现在上海石油化工研究院工作, 主要研究方向为过程监控和智能控制等。

蒋慰孙 见本刊 1999 年第 2 期第 168 页。