

一种基于 EEMD 的过程数据混合去噪方法^{*}

陈文驰, 刘 飞

(江南大学 信息技术系 轻工过程先进控制教育部重点实验室, 江苏 无锡 214122)

摘 要: 针对化工过程数据的特点, 提出一种基于集成经验模式分解(EEMD)滤波的过程数据混合去噪方法, 以新秩一阶差分法抑制数据粗差干扰, 以 EEMD 分解抑制脉冲干扰, 分层滤波消除噪声成分。与传统的滤波方法相比, 基于 EEMD 的混合滤波方法无须预先确定滤波器参数, 是一种完全的数据驱动型方法, 具有较好的自适应能力。仿真实验结果表明, 对过程数据的滤波预处理可以增强对异常突变数据的检测处理, 提高故障检测效果。

关键词: 粗差; 经验模式分解; 过程数据处理; 一阶差分

中图分类号: TP393.04 **文献标志码:** A **文章编号:** 1001-3695(2012)04-1368-03

doi:10.3969/j.issn.1001-3695.2012.04.046

Hybrid process data denoising method based on EEMD

CHEN Wen-chi, LIU Fei

(Key Laboratory of Advanced Process Control for Light Industry (Ministry of Education), Dept. of Information Technology, Jiangnan University, Wuxi Jiangsu 214122, China)

Abstract: In view of the features of chemical process data, this paper presented a novel integrated EEMD preprocessing method based on the principles of EMD denoising and first order differential. This method designed a first order differential method using new rank as the pre-filter process unit to reduce the effects of gross errors, and used a denoising scheme based on EEMD method to suppress pulse interference and remove white noise from the signal. Compared with traditional filtering, the hybrid EEMD filtering method did not need to define the coefficients of filter, so it was fully data-driven and adaptive. The simulation and experimental results demonstrate the effectiveness of this method in gross error elimination and fault detection.

Key words: gross error; empirical mode decomposition(EMD); process data processing; first order differential

在工业过程中, 由于仪表和执行器的噪声、输入扰动及传感器故障等原因往往使采集到的数据被噪声和粗差所干扰, 而数据质量对于后续的过程监测、在线优化以及各种控制算法的有效实施都具有重要影响, 所以进行过程数据预处理是非常必要的。近年来, 有学者将小波分析等先进的信号处理工具引入化工过程数据处理中^[1], 以完成对过程数据的去噪处理。常用的基于阈值滤波的小波降噪法都假设噪声干扰为白噪声, 当干扰分布不再是高斯分布或者有粗差干扰时, 阈值滤波就受到了限制。Nounou 等人^[2]给出了在线的不依赖于模型的能够剔除粗差的过程数据滤波方法。为提高滤波效果, 小波滤波法应根据具体问题选择合适的小波基, 许多应用中小波基选择必须依靠过程的先验知识, 在工程实际中针对不同的工程数据选择合适的基函数是一个困难的问题。近来, 一种新的信号分解算法——经验模式分解(EMD)由 Huang 等人提出, 其被认为是非线性、非平稳数据处理方法的新进展。EMD 可以从数据中自适应地得到基函数, 克服了小波变换中要选取合适的小波基的困难。本文针对化工过程数据的降噪预处理问题, 提出一种基于集成经验模式分解^[3](ensemble empirical mode decomposition, EEMD)的过程数据降噪方法, 并将其用于对吸附分离过程故障检测的数据预处理中。

1 基于秩的小样本数据粗差抑制

在实际的工业过程建模中, 测量数据不可避免地会有不确

定的干扰信息存在, 使得数据中往往包含一些粗差(异常点), 因此需要用某种方法来消除或减弱异常点的影响, 以利于后续的进一步降噪处理。常规的滑动中值滤波用于粗差抑制时, 虽然计算简单, 但是容易造成有用信号细节的损失。差分方法是一种常用的粗差处理方法, 其判别粗大误差的基础在于: 采样速率较高时, 邻近样本值的差值应较小, 即 $x(i-1) - x(i-2) \approx x(i) - x(i-1) \approx x(i+1) - x(i)$, 此时 $x(i)$ 的估值 $\hat{x}(i)$ 应满足 $\hat{x}(i) = x(i-1) + [x(i-1) - x(i-2)]$ 。当 $\hat{x}(i)$ 与 $x(i)$ 之差的绝对值超过设定的阈值 Δ 时(阈值 Δ 通常是标准差的整数倍), 便认为 $x(i)$ 含有粗差, 并用 $\hat{x}(i)$ 代替它。一般的差分方法消除粗差时采用的阈值是不抗差的, 利用不抗差的估计处理被污染的数据是不可靠的。于是便考虑利用分位数法, 用中位数代替平均值, 四分位数离散度作为方差的度量。由于分位数法利用了新秩排序, 含有粗差的样本值肯定是出现在数据的两端, 对中位数和四分位数离散度的影响极小, 因此分位数法具有良好的抗差性能。但是如果用分位数法直接处理原始数据, 进行粗差判别将存在两个缺陷: a) 有用信号会将弱小误差淹没, 使其不能被剔除; b) 用于稳定数据离散度的幂变换方法对数据的压缩严重, 使含较小粗差的数据被正常数据淹没, 不利于粗差检出。

基于上述原因, 对它们进行改进, 由于一阶差分所得数据离散度稳定, 粗差被扩散, 可采用新秩一阶差分分位数判

收稿日期: 2011-08-31; 修回日期: 2011-10-12 基金项目: 江苏省基础研究计划(自然科学基金)资助项目(BK2009068)

作者简介: 陈文驰(1978-), 男, 江苏无锡人, 博士研究生, 主要研究方向为工业系统的监测与诊断等(cwc1978314@163.com); 刘飞(1965-), 男, 安徽宣城人, 教授, 博士, 主要研究方向为工业过程监控和控制理论。

别^[4]:

a) 将数据样本 $x_1, x_2, \dots, x_n$ 按从小到大的顺序重新排序 $x_{(1)}, x_{(2)}, \dots, x_{(n)}$ 。

b) 将相邻数据差分, 进而形成新序列 $y(i) = x(i+1) - x(i)$, 将 $y(i)$ 按从大到小排序, 求得中位值 M 和上下四分位数 F_s 和 F_x , 求得平均值为 M 及四分位数离散度为 $\Delta = F_s - F_x$ 。

c) 如果新序两端的数据与中位值的差值大于 $k\Delta$ (k 根据样本的大小进行选值), 则认为该数据含有粗差, 并用其估值代替它, 这样保证样本的个数及平滑性。

2 基于 EEMD 过程数据去噪处理

2.1 EEMD 算法简介

EMD 方法认为任何动态信号都是由原始信号各阶固有模式函数 (intrinsic mode function, IMF) 合成的, 因此可以以某种方法将原始信号中所包含的各阶 IMF 分解出来, 每一阶 IMF 都反映原始信号的动态特性。IMF 分量必须满足两个条件: a) 其极点数和零点数相同 (或最多相差一个); b) 其上下包络线关于时间轴局部对称。EMD 方法分解的基本原理是通过不断剔除信号的极小值和极大值连接的上下包络线的均值, 将原始信号 $x(t)$ 分解为

$$x(t) = \sum_{j=1}^N \text{IMF}_j(t) + r_N \quad (1)$$

其中: r_N 为残余函数, 代表信号的平均趋势, 而各 IMF 分量 $\text{IMF}_1(t), \text{IMF}_2(t), \dots, \text{IMF}_N(t)$ 分别包含了信号从高到低不同频率段的成分。对含有噪声的过程数据进行 EMD 分解, 可以将过程数据特征在不同尺度上展现出来。这种多分辨率分解过程可以解释为以信号极值特征尺度为度量的时空滤波过程, 分解得到的各 IMF 分量都反映了信号的特征尺度, 代表着信号的内在模式特征。具体而言, 每一个 IMF 在时域上表现为某一尺度范围的模态, 并且彼此之间没有模态混叠。从频域上, 表现为从高频到低频的层层过滤。

在应用 EMD 降噪过程中, 当信号中存在异常事件 (如脉冲干扰、间歇性成分等) 时, 存在模式混叠现象。从前人的研究来看, 模式混叠的产生与极值点选择有关, 利用极值点信息通过样条插值生成的包络为异常事件的局部包络与真实信号包络的组合, 经该包络计算均值, 筛选出的 IMF 就包含信号固有模式和异常事件模式, 即模式混叠现象。在化工过程生产中, 由于仪器仪表的突然波动, 或装置条件的突然变化, 会使测量数据包含粗差等显著性误差。本文尝试在此基础上引入一种基于 EEMD 的降噪方法, 该方法引入了正态分布白噪声在 EMD 中具有的二阶时间尺度分解特性及不同白噪声序列对应 IMF 之间的无关性^[5,6], 一方面为分析信号提供了均匀分布的分解尺度, 另一方面添加的白噪声平滑了脉冲干扰等异常事件, 使异常事件模式在 EMD 分解过程中混入到白噪声模式中, 这在很大程度上抑制了异常事件模式和信号扰动固有模式的混叠, 更好地凸显真实信号。

2.2 基于 EEMD 的过程数据滤波算法

通常在实际的化工生产过程中, 需要在线地测量过程变量, 将其作为过程建模、操作与控制以及生产管理决策分析的直接依据。然而, 由于测量过程往往会受到各种不可预测因素的影响, 导致测量数据不可避免地包含误差, 影响过程信息处理和分析的效果。因此, 对现场采集来的测量数据在使用前进

行去噪处理是必要的。为实现对过程数据的在线 EEMD 去噪处理, 必须使用对短数据的局部 EMD 滤波^[7]处理。步骤如下:

a) 向被分析信号 $x(t)$ 中加入等长度的正态分布白噪声 (标准差取被分析信号标准差的 0.1 ~ 0.4 倍), 并对加入白噪声后信号进行归一化处理。

b) 应用 EMD 对加入白噪声后的信号进行分解, 初始化, $r(t) = x(t), i = 1$ 。

c) 如果余量 $r(t)$ 为单调函数或幅值小于设定阈值, 则算法停止, 否则转 d)。

d) 提取第 i 个 IMF 分量:

(a) 初始化 $h(t) = r(t)$;

(b) 计算 $h(t)$ 的局部极大值和极小值;

(c) 利用三次样条插值函数计算极大值包络线 $e_{\max}(t)$ 和极小值包络线 $e_{\min}(t)$;

(d) 计算 $e_{\max}(t)$ 和 $e_{\min}(t)$ 的平均值

$$m(t) = [e_{\min}(t) + e_{\max}(t)]/2$$

计算 $h(t) = h(t) - w(t)m(t), \delta(t) = \left| \frac{m(t)}{a(t)} \right|, a(t) = \frac{e_{\max}(t) - e_{\min}(t)}{2}$ 。设定三个门限值 $\theta_1, \theta_2, \alpha$, 默认值为 $\theta_1 =$

0.05, $\theta_2 = 0.5, \alpha = 0.95$ 。规定满足 $\delta(t) < \theta_1$ 的时刻个数与全部持续时间之比至少 $1 - \alpha$ 成立, 且对于每个时刻 t 有 $\delta(t) < \theta_2$ 。若满足此 IMF 中止条件, 转 e), 否则转 (b)。

e) 第 i 个 IMF 分量 $c_i(t) = h(t), i = i + 1$ 。

f) $r(t) = r(t) - c_i(t)$ 。

g) 重复上述 a) ~ f) 步骤 n 次, 每次添加不同的白噪声序列, 总共进行 n 次分解, 将每次分解得到的 IMF 对应集成平均, 当 n 足够大, 对应添加白噪声的 IMF 的和将趋于 0, 进而得到分析信号的固有模式函数组合为

$$x(t) = \sum_{i=1}^N \sum_{j=1}^n \text{IMF}_{j,i}(t) \quad i = 1, \dots, N; j = 1, \dots, n \quad (2)$$

3 对吸附分离过程数据的预处理

本文的仿真实验数据来自某石化厂的吸附分离过程。整个过程通过对塔釜、受槽液位、塔顶压力、汽化率、燃料油、燃料气压力、塔顶及各板温度、灵敏板温差等 15 个主要过程变量的监控, 实现安全、稳定、优质生产。该过程同其他工段的生产过程, 采用计算机分布式控制系统。

以吸附分离设备中某一段塔顶压力信号为例, 如图 1(a) 所示, 该信号包含大量随机噪声和一些粗差干扰, 这些噪声如果不处理, 会给后续的过程监控和故障诊断系统等造成较大影响。使用本文混合滤波方法对压力信号进行去噪处理, 先以基于差分方法的粗差抑制方法消除粗差干扰, 再以 EEMD 滤波消除随机噪声。处理结果如图 1(b) 所示, 噪声大大减少的同时, 还保留了原始信号的一些局部特征, 为后续系统的处理提供了较为真实的数据。

本过程有 15 个变量, 选择正常运行状态下的 200 个采样点以 PCA 法建立主元模型, 经分析因为前四个得分向量已经可以解释超过 95% 的数据信息, 所以建立只有四个主元的主元模型。另外选择 200 组含较多粗差和噪声的数据作为检验数据, 以主元分析法检测异常情况。下面分别给出用传统 PCA 方法和本文基于 EEMD 混合滤波的 PCA 方法在引入均值偏移

故障后所得到的 Hotelling T^2 图,如图 2、3 所示。

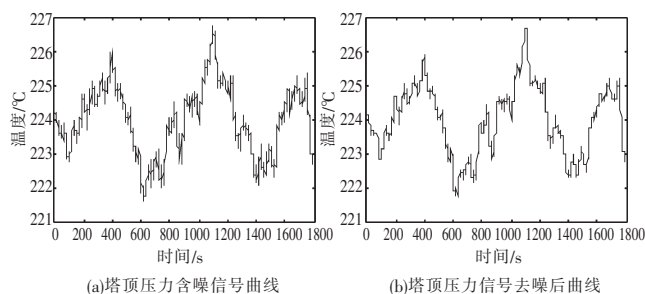


图1 塔顶压力信号去噪实例

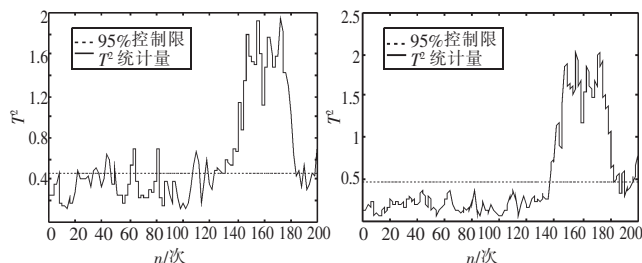


图2 传统主元分析 T^2 图

图3 基于EEMD混合滤波的主元分析 T^2 图

通过比较可以看出,经过滤波预处理后的正常过程数据波动明显变小,且在系统出现故障时, T^2 统计量超调更加明显,效果明显好于传统 PCA 方法的监控效果。本文的吸附分离过程是复杂的化工过程,过程数据突变性强,是典型的非平稳信号。由于传感器失灵、强电磁干扰以及设备故障等因素,过程数据中存在各种时变特征,如白噪声、阶跃、尖峰冲激以及趋势变化等,而且具有不可预测性、随机性,严重影响模型计算的精度,从而影响先进过程控制的在线应用效果。传统的基于主元分析的过程数据降维方法对数据误差较敏感,粗差和噪声的存在使过程主元发生剧烈波动。经过本文的数据滤波预处理,过程主元的异常波动得到平滑,而其突变特征和波形得到保持,提高了故障检测的准确性。传统主元分析的 T^2 图中第 60 和 80 采样点处含有超过控制限但却是正常工况数据的点,这些虚警点会引起系统的误报警,监控效果不理想。经

过基于 EEMD 混合滤波预处理后,在 T^2 图上消除了超过控制限的虚警点,这在极大程度上减少了系统的误报警,并能有效地检测出故障的发生。

4 结束语

实际工业生产过程中的过程数据不可避免地会有时变、不确定、不稳定、噪声和干扰等,这些不确定的干扰噪声会对正常的过程建模、数据分析操作、优化、故障检测等工作产生影响。因此,在进行此类工作前,必须对过程数据进行预处理。本文说明了如何用基于 EEMD 混合滤波对过程数据预处理的方法。基于一阶差分的粗差剔除方法在有效地抑制粗差干扰的同时保持了原始信号的局部特征,EEMD 分解滤波可以从数据中自适应地得到基函数,无须依靠过程的先验知识,可以有效滤除随机噪声。通过实例仿真可以看出,传统主元分析用于故障检测时,出现了很多误报警,而经过数据预处理的主元分析能很好地克服这个缺点,改善故障检测的效果。

参考文献:

(上接第 1358 页)此对液压站生产过程提出了许多改进措施并取得明显成效。由于机床的液压系统有诸多相似之处,可将此故障分析应用于其他型号的机床。从本文的分析过程中不难看出,故障数据对机床的可靠性分析和改进有重要作用,然而许多主机厂家缺少数据的积累,致使其不能全面彻底地推进可靠性工作。因此为了提高产品可靠性,企业在今后的产品维修过程中要重视收集故障数据,并在收集过程中要保证数据的全面性和完整性。

参考文献:

- [1] 周汝胜,焦宗夏,王少萍,等. 液压系统故障诊断技术的研究现状与发展趋势[J]. 机械工程学报,2006,42(9):6-13.
- [2] 王积伟,章宏甲,黄谊. 液压传动[M]. 北京:机械工业出版社,2010:23-42.
- [3] 明廷涛,张永祥,张西勇,等. 基于非线性观测器的液压伺服系统故障诊断方法研究[J]. 中国机械工程,2008,19(13):1527-1530.
- [4] 苏春,王金全,许映秋,等. 基于蒙特卡洛仿真的液压系统动态可靠性[J]. 东南大学学报:自然科学版,2006,36(3):370-373.
- [5] 周汝胜,焦宗夏,王少萍,等. 基于专家系统的导弹发射车液压系统故障诊断[J]. 航空学报,2008,29(1):197-203.
- [6] 贺湘宇,何清华,谢习华,等. 基于偏最小二乘回归的挖掘机液压

- 系统故障诊断[J]. 中南大学学报:自然科学版,2007,38(6):1152-1156.
- [7] 刘品,刘岚岚. 可靠性工程基础[M]. 北京:中国计量出版社,2009.
- [8] XIE Lei, ZHANG Jian-ming, WANG Shu-qing. Sensor fault diagnosis with statistical signal reconstruction approach[J]. Journal of Chemical Industry and Engineering,2006,57(10):2343-2348.
- [9] DEKKER A J D, SIJBERS J. Implications of the Rician distribution for fMRI generalized likelihood ratio tests[J]. Magnetic Resonance Imaging,2005,23(9):953-959.
- [10] 周正伐. 可靠性工程基础[M]. 北京:中国宇航出版社,2009.
- [11] YANG Hai-lan, CAI Yan, BAO Ye-feng, et al. Analysis and application of partial least square regression in arc welding process[J]. Journal of Central South University of Technology,2005,12(4):453-458.
- [12] TAN Hong-zhou, SEPEHRI N. Parametric fault diagnosis for electro-hydraulic cylinder drive units[J]. IEEE Trans on Industrial Electronics,2002,49(1):96-106.
- [13] CHEN Ling-xin, WANG Hai-long, MA Ji-ping, et al. Fabrication and characterization of a multi-stage electro-osmotic pump for liquid delivery[J]. Sensors and Actuators A: Chemical,2005,104(1):117-123.
- [14] 杨晓川,谢庆华,何军,等. 基于故障树的模糊故障诊断方法[J]. 同济大学学报:自然科学版,2001,29(9):1058-1060.