

# 小波变换的数据预处理在软测量建模中的应用

窦伟<sup>1</sup>, 张滢<sup>1</sup>, 薄翠梅<sup>1</sup>, 蒋楠<sup>2</sup>

DOU Wei<sup>1</sup>, ZHANG Shi<sup>1</sup>, BO Cui-mei<sup>1</sup>, JIANG Nan<sup>2</sup>

1. 南京工业大学 自动化学院, 南京 210009

2. 西安市环境保护研究所, 西安 710054

1. College of Automation, Nanjing University of Technology, Nanjing 210009, China

2. Xi'an Munciple Environmental Protection Research Institute, Xi'an 710054, China

**DOU Wei, ZHANG Shi, BO Cui-mei, et al. Data pre-processing for soft-sensor modeling in chemical process using wavelet transform. Computer Engineering and Applications, 2009, 45(14): 201-203.**

**Abstract:** Data-driven soft-sensing is based on the large number of industrial data, so the data pre-processing is an important task for soft-sensing modeling. In this paper, a new method is used to process the gross error result in noise. That is a compromising algorithm between soft-threshold and hard-threshold of wavelet coefficient through adding a weight for threshold function. Simulation result and industrial application show that this method proposed is feasible and effective.

**Key words:** wavelet transform; soft-hard threshold; soft-sensing

**摘要:** 数据驱动的软测量建模的基础是大量可靠而准确的工业现场数据, 因此数据预处理是软测量建模的重要任务。提出将小波系数软硬阈值折中方法应用于去除噪声干扰所产生的随机误差, 即通过增加一个权值的应用, 以综合软阈值和硬阈值两种方法的优点。仿真实验和工业实例应用证明了该方法的可行性和有效性。

**关键词:** 小波变换; 软硬阈值; 软测量

**DOI:** 10.3778/j.issn.1002-8331.2009.14.062 **文章编号:** 1002-8331(2009)14-0201-03 **文献标识码:** A **中图分类号:** TP274

## 1 引言

软测量技术已经在流程工业中得到了很好的应用, 其核心是通过一些能够测量或者便于测量的数据, 构成某种数学关系通过计算机软件来推断和估计那些难于或者暂时无法测量的变量, 以满足工业生产的要求。

在软测量建模所使用的方法中, 目前较多使用的是数据驱动型的建模方法, 因此数据样本数量与质量的好坏对于建模效果起着至关重要的作用。一般来说, 进行建模对象的样本数量是有限的, 因此在建模前进行必要的预处理以使这有限的准确、完整地描述对象, 成为提高建模精度的一个关键步骤。数据预处理中很重要的任务即是进行数据误差处理。

根据化工过程中测量数据性质及误差产生的原因, 可将误差分为随机误差和过失误差。随机误差主要受随机因素的影响, 如操作过程的微小波动或检测信号的噪声等, 其中工业现场的有色噪声又是产生随机误差的最主要原因; 过失误差主要包括仪表的系统偏差(如堵塞、校正不准、基准漂移等)、测量设备失灵等以及人为某些人为因素有关<sup>[1]</sup>。随机误差的处理, 尤其是噪声干扰所引起的随机误差一直以来就是数据预处理中的

难点和重点<sup>[2]</sup>, 目前使用比较多的随机误差处理方法是滑动时间窗平均滤波方法, 虽然该算法实施简便, 但其精度低可靠性差, 尤其是在噪声干扰较大且为有色噪声的情况下无法对测量数据的过失误差进行有效去除。

主要研究的是将小波系数软硬阈值折中方法应用于软测量建模中随机误差的处理, 以去除噪声对测量数据的影响。

## 2 小波变换降噪及小波系数的软硬阈值折中方法介绍

### 2.1 小波变换降噪的基本思想

小波变换是空间(时间)和频率的局部变换, 因而能有效地从信号中提取信息。通过伸缩和平移等运算功能可对函数或信号进行多尺度的细化分析, 解决了 Fourier 变换不能解决的许多困难问题。随着小波分析理论研究的不断深入, 利用小波变换进行信号去噪及重构开始被广为关注。小波变换能同时利用信号与噪声在时域和频域内的差别, 可实现更为有效的信噪分离, 从而获得较为理想的除噪效果; 同时小波变换具有低熵性、多分辨率、去相关性等优点<sup>[3]</sup>。

设有如下的观测信号

**基金项目:** 江苏省高校自然科学基金资助项目(Natural Science Foundation for Colleges and Universities in Jiangsu Province, under Grant No. 07KJB510042); 扬子石化公司资助项目(The Projects Funded by Yangtze Petrochemical Company, under Grant No.04JSNJY110019)。

**作者简介:** 窦伟(1982-), 男, 硕士研究生, 主要研究方向: 系统建模、仿真与优化等; 张滢(1952-), 女, 博士, 教授, 主要研究方向: 先进控制技术及其工业应用; 薄翠梅(1973-), 女, 博士生, 讲师, 主要研究方向: 非线性系统的故障诊断及主动容错控制; 蒋楠(1969-), 男, 工程师, 主要研究方向: 自动控制技术的应用。

**收稿日期:** 2008-03-17 **修回日期:** 2008-06-06

$$f(t)=s(t)+n(t) \quad (1)$$

其中  $s(t)$  为原始信号,  $n(t)$  为方差为  $\sigma^2$  的 Gaussian 白噪声, 服从  $N(0, \sigma^2)$ 。降噪处理的目的就是尽量从观测信号  $f(t)$  中把有用的信号  $s(t)$  提取出来。

对于一维信号  $f(t)$  来说, 首先对其进行离散采样, 等到  $N$  点离散信号  $f(n)$ ,  $n=0, 1, \dots, N-1$ , 其小波变换为:

$$Wf(j, k)=2^{-j/2} \sum_{n=0}^{N-1} f(n)\phi(2^{-j}n-k) \quad (2)$$

$Wf(j, k)$  即为小波系数, 在实际应用中, 直接利用式(2)计算比较繁琐, 因此需要借助双尺度方程, 可得到小波的递归实现方式:

$$Sf(j+1, k)=Sf(j, k) \times h(j, k) \quad (3)$$

$$Wf(j+1, k)=Sf(j, k) \times g(j, k)$$

其中  $h$  和  $g$  分别是对应于尺度函数  $\varphi(x)$  和小波函数  $\phi(x)$  的低通和高通滤波器,  $Sf(0, k)$  为原始信号  $f(k)$ ,  $Sf(j, k)$  为尺度系数,  $Wf(j, k)$  为小波系数。相应地, 小波变换重构公式为:

$$Sf(j-1, k)=Sf(j, k) \times \tilde{h}(j, k) + Wf(j, k) \times \tilde{g}(j, k) \quad (4)$$

为了便于表示, 小波系数  $Wf(j, k)$  简记为  $w_{j,k}$ , 对观测信号  $f(x)=s(k)+n(k)$  作离散小波变换之后, 由小波变换的线性性质可知, 分解得到的小波系数  $w_{j,k}$  仍然由两部分组成, 一部分是  $s(k)$  对应的小波系数  $Wf(j, k)$ , 记为  $u_{j,k}$ , 另一部分是  $n(k)$  对应的小波系数  $Wn(j, k)$ , 记为  $v_{j,k}$  [4]。

算法基本的步骤可表达为:

(1) 先对含噪声信号  $f(k)$  作小波变换, 得到一组小波系数  $w_{j,k}$ ;

(2) 通过对  $w_{j,k}$  进行阈值处理, 得到估计小波系数  $\hat{w}_{j,k}$ , 使得  $\|\hat{w}_{j,k}-u_{j,k}\|$  尽可能小;

(3) 利用  $\hat{w}_{j,k}$  进行小波重构, 得到估计信号  $\hat{f}(k)$ , 即为去噪之后的信号。

## 2.2 小波系数的软硬阈值折中方法

根据 Donoho 提出的统一阈值方法, 对信号进行小波分解, 分解后得到近似系数和细节系数, 再对细节系数作用阈值抑制噪声, 重建信号达到降噪的目的。目前已有将软阈值或硬阈值方法应用于仿真研究或实际应用中 [4]。

但是这种统一阈值方法在实际的应用中效果欠理想; 硬阈值尽管在使用中计算简单方便, 但得到的估计小波系数值 (EWC) 连续性差, 可能引起重构信号的振荡; 而软阈值方法能够根据噪声变化的情况, 选择不通的阈值, 因此得到的估计小波系数连续性好, 但是当小波系数较大时, EWC 与原来的小波系数有固定的偏差, 也会给重构的信号带来不可避免的误差。针对此问题, 提出将软硬阈值折中方法应用于随机误差处理 [5-6]。

与通常的软阈值或硬阈值估计不同的是, 在表达式中引入了权值  $\alpha$  以综合这两种方法的优点, 可将其的估计表达式定义为:

$$\hat{w}_{j,k} = \begin{cases} \text{sign}(w_{j,k}) \cdot (|w_{j,k}| - \alpha\lambda), & |w_{j,k}| \geq \lambda \\ 0, & |w_{j,k}| < \lambda \end{cases} \quad 0 \leq \alpha \leq 1 \quad (5)$$

特别的当  $\alpha$  分别取 0 和 1 时, 上式即成为硬阈值和软阈值估计方法。图 1 中的实线和点线分别表示的是硬阈值和软阈值方法的阈值函数, 虚线表示的是软硬阈值折中方法的阈值函数。对于一般的  $0 < \alpha < 1$  来讲, 该方法估计出来的数据  $\hat{w}_{j,k}$  的大小介于软、硬方法之间, 故称之为软硬阈值折中方法。这种处理的方法

尽管简单, 但是去噪的效果非常的好。根据不同的使用情况进行  $\alpha$  的选择, 在本文的研究中取  $\alpha$  为 0.6

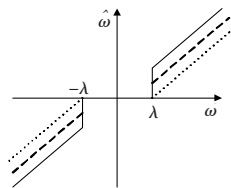
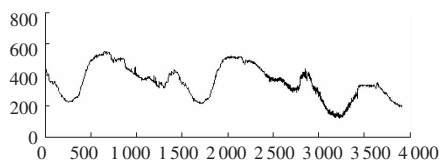


图 1 阈值函数

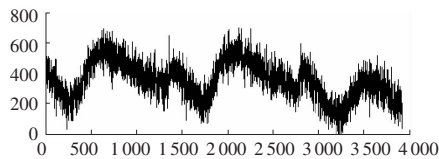
## 3 仿真研究

采用如下的仿真算例来检验小波系数软硬阈值折中方法的去噪效果。

生成如图 2(a) 所示的信号作为仿真研究的输入信号, 对其加入白噪声, 使信噪比 (SNR) 为 14 左右, 得到如图 2(b) 所示的信号。



(a) 原始输入信号



(b) 加入噪声后的输入信号

图 2 仿真输入信号

Daubechies 小波由于其对非平稳信号的灵敏性得到了广泛的应用。根据 Daubechies 小波的正交、时频紧支撑、高正规性和具有 Mallat 快速算法等特点, 采用不同  $N$  值的 Daubechies 小波, 可以得到不同的处理效果。  $N$  越大, 其滤波器的长度越长, 滤波性能越好, 但信号处理的时延也越长, 即时域定位性变差。

选用 Daubechies8 小波作为小波函数进行含噪声信号的去噪和重构, 为便于验证软硬阈值折中方法的去噪效果, 选用目前比较常用的软阈值方法和滑动时间窗平均滤波方法对图 2 所示的仿真模拟信号, 进行降噪处理。

处理的结果如图 3 所示, 其中的 (a), (c), (e) 三个图分别表示的是软硬阈值折中方法、软阈值方法和滑动时间窗平均滤波方法在去除噪声后得到的原始信号曲线。右边与之对应的 (b), (d), (f) 三个图是降噪处理后得到的信号与原始输入信号的偏差。通过分析偏差的大小可以发现, 三种方法中软硬阈值折中方法有着最小的偏差, 而滑动时间窗平均方法在降噪处理的效果上不明显。

## 4 工业应用实例

为进一步地检验小波系数软硬阈值折中方法在随机误差处理方面的效果。根据本研究所对某化工企业丁二烯生产装置应用基于支持向量机方法建立软测量模型的科研成果 [7], 将该方法应用于软测量建模前数据预处理。

该丁二烯生产装置的 DA-1107 塔是丁二烯产品分离塔, 以塔顶丁二烯成品丁二烯-1, 3 的纯度变量 S1136 为建模的

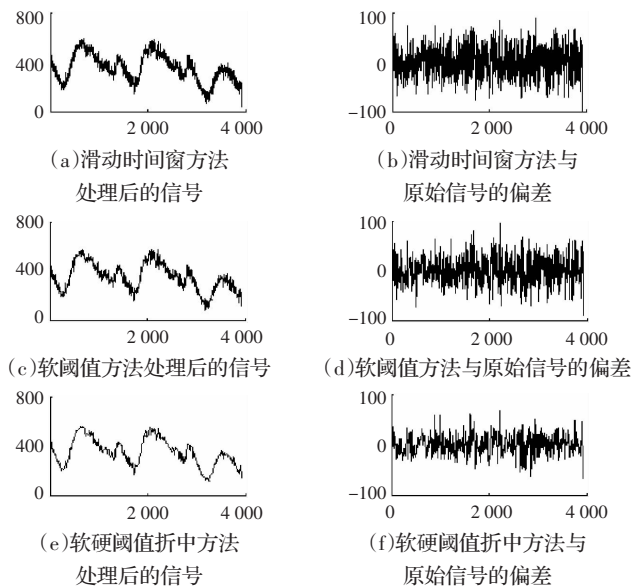


图3 各种降噪处理方法的效果对比

目标。由于该纯度直接反映丁二烯-1,3的产品质量,因此对工厂的经济效益有着重要的影响,对软测量模型的精度和可靠性要求比较高。通过对工艺进行分析和变量的筛选,其的值主要受到 DA1106 塔进料流量(F1122)、DA1107 塔塔釜排放量(F1138)、回流量(F1139)、塔顶压力(P1120)、塔顶温度(T1500-30)和塔釜温度(T1300-6)等 6 个变量的影响<sup>[8]</sup>,因此选用这 6 个变量作为输入值,成分值 S1136 作为输出值,并且以成分分析化验的结果作为标准。

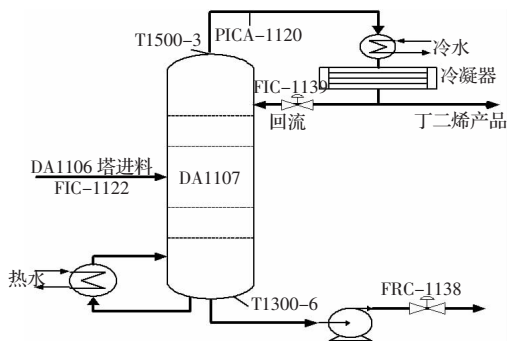


图4 丁二烯 DA1107 塔工艺流程

应用小波系数的软硬阈值折中方法进行建模数据预处理,以最大限度地消除噪声引起的随机误差的影响并尽可能保留信号的突变特征,删掉不必要的干扰,选取与上面仿真实验相同的 Daubechies8 小波,三层分解进行去噪处理。处理后的数据用来训练本实验室已建立的支持向量机软测量模型,模型的输出结果如图 5 所示,模型的外推精度和拟合精度有了不同程度的提高,尤其是对外推精度的效果比较明显。

为便于评价不同模型的精度,选取相对均方根误差  $R_{RMSE}$  作为的评价准则,如式(6)所示。其中  $N$  表示建模数据的序列长度,  $y_i, y_i^m$  分别表示  $i$  时刻分析数据值和模型的输出值。表 1 列出了应用小波系数软硬阈值折中方法预处理后得到的模型与本实验室应用滑动时间窗平均滤波方法预处理后得到的模型的相对均方根误差比较,可以看出,应用小波系数软

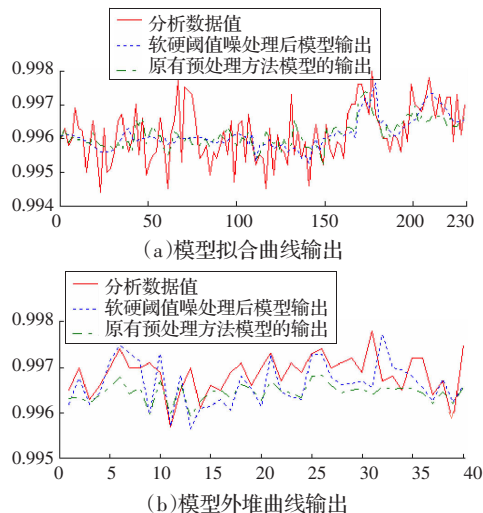


图5 不同数据预处理方法拟合精度和外推精度比较

表1 不同预处理方法对模型精度的影响

	拟合精度	外推精度
软硬阈值折中方法	5.100 7e-004	4.218 2e-004
原有预处理方法	6.625 5e-004	5.596 5e-004

硬阈值折中方法预处理后得到的模型有着更好的拟合精度和外推精度。

$$R_{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N \left( \frac{y_i - \hat{y}_i}{y_i} \right)^2} \quad (6)$$

### 5 结论

小波系数软硬阈值折中方法的采用,能够有效地去除噪声引起的随机误差。仿真研究和工业实例应用表明针对相同的建模方法,经过该方法预处理后数据的建模效果优于目前较多采用的软阈值方法和滑动时间窗平均滤波方法。但是由于建模方法本身的不足以及间隔 8 小时的分析数据,致使软测量的计算值存在着较大的滞后,这是在今后的研究工作中需要进一步解决的问题。

### 参考文献:

- [1] 黄德先,叶心宇,竺建敏,等.化工过程先进控制[M].北京:化学工业出版社,2006:57-67.
- [2] 罗健旭,常青.软测量技术的数据预处理方法研究[J].控制工程,2006(4):298-300.
- [3] Donoho D. De-noising by soft-thresholding[J]. IEEE Trans on Information Theory, 1995.
- [4] Sardy S, Tseng P, Bruce A. Robust wavelet denoising[J]. IEEE Trans on Signal Processing, 2001,49(6): 1146-1152.
- [5] 徐晨,赵瑞珍,甘小冰.小波分析·应用方法[M].北京:科学出版社,2004:104-116.
- [6] 周静,陈允平,周策,等.小波系数软硬阈值折中方法在故障定位消噪中的应用[J].电力系统自动化,2005,29(1):65-66.
- [7] 崔永超,张混,王永华.基于支持向量机的软测量建模方法的应用[J].南京工业大学学报:自然科学版,2007(3).
- [8] 薄翠梅,张混.基于神经网络的软测量技术在精馏塔上的应用[J].过程工程学报,2003,3(4):371-375.