Journal of Vibration, Measurement & Diagnosis

doi:10.16450/j.cnki.issn.1004-6801.2019.02.027

基于样本熵的改进小波包阈值去噪算法

向北平,周建,倪磊,艾攀华

(西南科技大学制造过程测试技术教育部重点实验室 绵阳,621000)

摘要为了更好地消除噪声对被测振动信号的干扰,分析了样本熵算法与噪声的关系,提出了一种基于样本熵的 改进小波包阈值去噪算法。在阈值函数方面,该方法利用样本熵作为特征参数,对含噪信号小波包系数的噪声分 布进行表征,且依据此特征参数值对阈值函数进行改进,使其能够根据信号的小波包系数受噪声影响的情况进行 自适应的调整;在阈值选取方面,定义去噪后信号与原始信号之差作为噪声信号的估计,利用样本熵作为判别依 据,选取使得噪声估计的样本熵值最大的阈值作为最优阈值。该方法与其他方法进行对比,结果表明,该方法能够 有效地去除噪声且更好地还原信号的频率特征,是一种更为优越的去噪算法。

关键词 样本熵;小波包系数;阈值去噪;振动信号;故障诊断 中图分类号 TN911.7;TH165⁺.3

引 言

长期以来,小波包滤波是对信号进行分析预处 理的主要工具,该方法对信号进行小波包变换,然后 利用小波包阈值函数对小波包系数进行阈值收缩处 理以达到去噪目的。然而由于含噪信号中的噪声分 布往往不均匀,传统的软、硬阈值方法对信号小波包 系数进行固定格式的阈值处理,无法很好地满足信 号去噪要求^[1]。另外,在阈值的选择方面,常用的 Heursure 阈值、Donoho 阈值等能够对信噪比较高 的信号实现噪声与信号的最优分离,而对于强噪声 信号,去噪效果并不理想^[2]。近年来,针对这两个问 题,许多学者进行了研究。

文献[3]提出了一种自适应对数小波阈值函数 去噪算法,结合自适应对数阈值函数对每一层小波 系数设置最优阈值,且应用于动压信号,增加了信号 信噪比且减少了计算时间,但其去噪算子形式依然 固定不变。文献[4]分析了传统软、硬阈值方法的局 限,对阈值函数进行改进,使其具有能量分布自适应 性,但信号小波包系数的能量分布并不能准确表达 小波包系数的噪声分布,因此该方法仍然存在一些 不足。文献[5]对 Donoho 阈值方法中的噪声标准 差估计方法进行改进,进行仿真且取得了良好的去 噪效果。Richman 等^[6]提出一种新的时间序列复杂 度表征参数即样本熵(sample entropy,简称 $S_{\rm E}$)算法,被广大学者所关注且近年来被常用于机械信号分析与故障诊断领域^[7]。

基于以上分析,笔者提出将样本熵与小波包阈 值去噪算法相结合,且将其应用于高速深沟球轴承 振动信号去噪分析,通过分析信号的小波包系数噪 声分布情况及其对应的样本熵值,使阈值去噪算子 具有噪声分布自适应性以达到最优去噪效果;同时 以噪声估计信号样本熵值为基准,提出了一种最优 阈值估计方法。仿真分析以及实验结果皆对所提方 法进行了验证。

1 样本熵算法及其噪声表征

1.1 样本熵的计算

设有长度为 N 的时间序列 $X_i = \{x_1, x_2, \dots, x_N\}$,其样本熵的计算方法如下:

1) 确定嵌入维数为 m, 对 X_i 的元素按顺序进行 排列, 即可得到一组维数为 m 的向量 $\{x_m(1), \dots, x_m(N-m+1)\}, 且$

 $X_m(i) = \{x(i), x(i+1), \dots, x(i+m-1)\}$

 $(1 \leqslant i \leqslant N - m + 1) \tag{1}$

2) 定义向量 $X_m(i) = X_m(j)$ 之间的间隔 $d[X_m(i), X_m(j)]$ 为两向量之间对应元素求差的绝 对值的最大值,即

^{*} 国家重大科学仪器设备开发专项基金资助项目(2013YQ13042902);西南科技大学博士研究基金资助项目(15zx7122) 收稿日期:2017-05-02;修回日期:2017-12-05

$$d \left[\boldsymbol{X}_{m}(i), \ \boldsymbol{X}_{m}(j) \right] =$$

$$\max_{k=0,\dots,m-1} (|x(i+k) - x(j+k)|)$$
 (2)

3) 对于固定的 $X_m(i)$,统计 $X_m(i)$ 与 $X_m(j)$ 之 间距离小于等于参数 r 的 j (1 $\leq j \leq N-m, j \neq i$)的 个数,且记为 B_i ,则当 1 $\leq i \leq N-m$ 时定义

$$B_i^m(r) = \frac{1}{N - m - 1} B_i \tag{3}$$

4) 定义 B^(m)(r)为

$$B^{(m)}(r) = \frac{1}{N-m} \sum_{i=1}^{N-m} B_i^m(r)$$
 (4)

5)将维数增加到 m+1,同样按照上述方法计算 $X_{m+1}(i)$ 与 $X_{m+1}(j)$ (1 $\leq j \leq N-m, j \neq i$)之间距离小 于等于 r的个数,且记为 A_i ,则有 $A_i^m(r)$ 定义为

$$A_{i}^{m}(r) = \frac{1}{N - m - 1} A_{i}$$
(5)

6) 定义 A^m(r)为

$$A^{m}(r) = \frac{1}{N-m} \sum_{i=1}^{N-m} A_{i}^{m}(r)$$
(6)

据上述分析可知,B^(m)(r)是两个序列在相似容限 r 下匹配 m 个点的概率,而 A^m(r)是两个序列匹 配 m+1 个点的概率。则该时间序列样本熵定义为

$$S_{\rm E}(m,r) = \lim_{N \to \infty} \left\{ -\ln\left[\frac{A^m(r)}{B^m(r)}\right] \right\}$$
(7)

实际信号中 N 无法趋近于无穷,因此可将样本 熵设为

$$S_{\rm E}(m,r,N) = -\ln\left[\frac{A^m(r)}{B^m(r)}\right] \tag{8}$$

上述算法中的嵌入维数 m 和相似容限 r 通常

取为m=1或2,r=0.1, $S_{td} \sim 0.25S_{td}$ (S_{td} 为原始数 据 $X_i = \{x_1, x_2, \dots, x_N\}$ 的标准差)。文中取m=2, $r=0.2S_{td}$ 。

1.2 样本熵与噪声

当信号受到噪声干扰后,其状态取值不确定性 增加,即信号无序程度与复杂程度增加。而样本熵 作为信号复杂度表征参数,当信号中噪声增加,样本 熵值也应增加。为验证以上分析,设定1kHz采样 率与1000总采样数的仿真信号 f(t) = y(t) + n $(t),式中n(t)为标准差 \sigma \in [0,3]的带限 500 Hz 高$ 斯白噪声信号。

 $y(t) = 5\sin(10\pi t)\sin(2\pi t) \qquad (0 \leqslant t \leqslant 1) \ (9)$

分析信号 f(t)的样本熵值在不同采样数下与 噪声标准差的关系,其结果如图 1(a)所示(图中 100 采样点为信号前 100 采样点,下同)。由于实际工程 信号中的噪声很少为单纯的白噪声,因此,利用反幂 律滤波器对功率谱密度分布均匀的白噪声n(t)上 色,选定反频谱指数为 1,此时理想数字滤波器的幅 度平方响应为 $1/f^1$ 。由于该滤波器处理后导致噪 声幅值衰减,因此,对加色后的噪声信号进行 10 倍 处理,得到有色噪声 $n_1(t)$ 及含噪仿真信号 $f_1(t)$ = $y(t)+n_1(t)$ 。设 $n_2(t)$ 为 1 kHz 采样率,1 000采样 数的幅值区间为[0,3]的带限 500Hz 周期性随机噪 声,得到含噪仿真信号 $f_2(t) = y(t) + n_2(t)$ 。同样分 析信号 $f_1(t)$ 与 $f_2(t)$ 的样本熵值随噪声大小的变化 关系,其结果分别如图 1(b),(c)所示。





由图 1(a)看出,信号的样本熵值与噪声标准差 成正相关,且在采样点相差巨大的情况下其样本熵 值变化依然相近。图 1(b),(c)较图 1(a)而言,样本 熵随噪声标准差变化有少许波动,熵值大小区间有 变化,但总体趋势仍呈正相关,且数据长度对其趋势 影响不大。这说明,虽然对仿真信号施加不同种类 的噪声,其样本熵变化曲线有所不同,但由于对信号 增加噪声的结果导致了信号的随机度与复杂度增 加,其结果仍然是样本熵增加。图 1 所示说明了样 本熵算法可用来表征时间序列的含(多种)噪声情况 且受数据长度影响较小。

2 改进小波包阈值去噪算法

2.1 传统小波包去噪及其不足

传统的阈值函数主要有软阈值函数和硬阈值函 数两种。软阈值函数

$$\eta(w,\lambda) = \begin{cases} \operatorname{sgn}(w) \left(\left| w \right| - \lambda \right) & \left(\left| w \right| \ge \lambda \right) \\ 0 & \left(\left| w \right| < \lambda \right) \end{cases}$$
(10)

硬阈值函数

$$\eta(w,\lambda) = \begin{cases} w & (|w| \ge \lambda) \\ 0 & (|w| < \lambda) \end{cases}$$
(11)

工程实践中常用的阈值函数还有形如文献[8] 中提出的一种介于软、硬阈值函数之间的改进阈值 函数

$$\eta(w,\lambda) = \begin{cases} \operatorname{sgn}(w) \left[|w| - \frac{2\lambda}{1 + \exp((|w| - \lambda))} \right] (|w| \ge \lambda) \\ 0 & (|w| < \lambda) \end{cases}$$

由以上函数式可知:硬阈值由于其函数不连续 性可能产生伪 Gibbs 现象且导致有用信号缺失;软 阈值函数虽然连续性好,但存在较大偏差^[9-10]。文 献[8]阈值函数虽进行了改进,但只是软硬阈值函数 的折中算法,通过下文分析可知,该阈值函数仅为本 研究提出的自适应阈值函数的一种取值情况(即当 调节参数 *s*=0.5 时)。在实际应用中,阈值函数的 选择并无固定标准,对于不同的信号选用不同的去 噪算子达到的去噪效果也不同^[11],因此,研究一种 能依据信号小波包系数的含噪情况自适应调整的新 阈值函数具有实际的工程意义。

2.2 自适应改进阈值函数

根据以上分析,笔者对传统阈值函数作改进,使 其不受限于固定去噪形式,得到的改进阈值函数为 $\eta(w,\lambda,s) =$

$$\begin{cases} \operatorname{sgn}(w) \left[|w| - \frac{\lambda}{\exp\left((|w| - \lambda) \frac{1 - s}{s}\right)} \right] (|w| \ge \lambda) \\ 0 & (|w| < \lambda) \end{cases} \end{cases}$$
(13)

其中:0 < s < 1为调节参数,当 $s \rightarrow 0$ 时,该阈值函数 接近为硬阈值函数;当 s = 1时,该阈值函数为软阈 值函数。

由于阈值函数偏硬时对大于阈值的小波包系数 保留较好,而偏软时对小波包系数压缩更大。因此, 对于含噪较多的小波包系数,阈值处理方式应偏软 即 *s* 较大;而受噪声影响小的小波包系数阈值函数 应偏硬,即 *s* 应较小。

由 1.2 节分析可知,样本熵能较好地反应时间 序列的噪声变化情况,且适用于分析短时间序列,因 此设定调节参数 s 的确定方法如下:

1) 对信号的小波包系数序列(w_1, w_2, \dots, w_n)

按顺序分割为n-l个相互之间最大重叠且长度均为l的子序列 (k_1,k_2,\dots,k_{n-l}) ,即 k_i 向后移动一位数据得到 k_{i+1} ;

2) 对上述 n-l 个子序列分别计算其样本熵 值,且将该值作为子序列中间数据点的样本熵值,即 可得到相应的样本熵序列 $\tilde{s} = \{\tilde{s}_1, \tilde{s}_2, \dots, \tilde{s}_{n-l}\}$ (左右 两端缺失的样本熵数据按边缘值进行延拓);

3)将该序列进行极值归一化后带入改进阈值 函数中,可使其具有小波包系数噪声分布自适应性。

2.3 阈值估计

阈值的作用是将信号与噪声系数进行分割,将 信号系数进行重构即可得到去噪后信号 $\tilde{y}(t)$,则原 始信号f(t)中噪声信号的估计为 $\tilde{n}(t) = f(t) \tilde{y}(t)$ 。当噪声去除越多, $\tilde{n}(t)$ 的样本熵越大;而当信 号被过扼杀,即 $\tilde{n}(t)$ 中含有部分有用的规律性信 号,此时 $\tilde{n}(t)$ 复杂度降低,其样本熵值也会相应减 少。据此笔者利用样本熵作为判据,选取使得噪声 信号的估计 $\tilde{n}(t)$ 的样本熵最大时的阈值作为最佳 阈值,认为此时噪声去除最为彻底且保留原始信号 最多,即阈值估计公式为

$$\lambda = \lambda \left\{ S_{\rm E} \left\{ \tilde{n}(t) \right\} \right\} \tag{14}$$

分析 1.2 节中仿真信号,对其加入标准差为 1 的 混合噪声,该混合噪声包含上文提到的的高斯白噪 声、有色噪声以及周期随机噪声,按上述方法进行阈 值估计,得到阈值与 $S_{\rm E}\{\hat{n}(t)\}$ 的关系如图 2 所示。



Fig. 2 Correlation between threshold and $S_{\rm E}(\tilde{n}(t))$

如图 2 所示,按本方法计算得到的阈值为 λ = 0.6,为进行对比,设定其他阈值分别为 λ =0.5, 0.7,利用文中自适应改进阈值函数(sym8 小波,分 解层数为 3 层)分别选定上述 3 个阈值进行去噪分 析,得到结果如图 3 所示。由图 3 可知,当阈值 λ = 0.5 时,阈值过小,信号毛刺较多,噪声去除不完全; 当阈值取 0.7 时,原始信号被过度压缩;而通过本方 法确定的阈值取得了更好的去噪效果,证明了该阈 值估计方法的有效性。

为了验证该去噪方法受噪声标准差的影响,同 样分析上述信号,分别加入不同标准差的混合噪声, 对其进行去噪,计算去噪前后的信噪比(signal to



图 3 不同阈值去噪效果对比



noise ratio,简称 SNR)与均方根误差(root mean square error,简称 RMSE),且将结果列于表 1。由表 1 可知,对不同噪声标准差情况下的含噪信号进行去噪,去噪后信号的 SNR 与 RMSE 变化幅度不大,证明该方法受噪声标准差影响较小,适用于对不同噪声尺度情况下的信号进行去噪。

表 1 不同噪声尺度下的去噪结果 Tab. 1 Denoising results of different noise scale

噪声 标准差	0.5	1.0	1.5	2.0	2.5	3.0
SNR	16.24/	12.01/	8.86/	5.86/	4.54/	2.38/
(去噪前/后)	31.87	29.65	28.7	27.53	27.48	26.95
RMSE	0.51/	1.03/	1.46/	1.94/	2.55/	2.93/
(去噪前/后)	0.07	0.08	0.08	0.09	0.10	0.11

3 实 验

3.1 振动实验装置与信号获取

为了对本算法进一步验证,搭建了轴承振动实 验平台如图 4 所示,该实验平台主要包括中间的小 型高速直流电机,两端测试轴承为分子泵中使用的 微小型高速深沟球轴承 QC0011286(轴承参数为: 内径为 4 mm,外径为 13 mm,轴承节圆直径 D 为 8.5 mm,滚珠直径 d 为 2.3 mm,滚珠个数 N 为 7 个,接触角 β 为 0°)以及固定于轴承外圈的加速度传 感器。测试时电机转速设定为 60 kr/min,采样率 为 20 kHz,采样时间为 0.1 s,采样时前置加上 10 kHz的低通抗混滤波。为模拟轴承外圈故障,在 轴承外圈内壁加工宽为 0.15 mm、深为 0.2 mm 的 横向沟槽,此时轴承基频为 f_r=1 kHz,轴承外圈故 障频率根据公式(15)可得

$$f_{\rm oc} = \frac{N}{2} \left[1 - \frac{d}{D} \cos\beta \right] f_r = 2 \ 552.9 \ \text{Hz} \quad (15)$$

采集到的轴承振动时域波形如图 5 所示。由图 5 可知,从该时域信号无法直观地得出任何轴承信 号特征。



图 4 轴承振动实验平台 Fig. 4 Bearing vibration experiment platform



Fig. 5 Bearing vibration time domain signal

3.2 去噪分析与对比

dB

对图 5 所示的信号选用 sym8 母小波进行四层 小波包分解,获得最大尺度上的小波包系数序列 w(i),为进行样本熵计算,对其分割为若干个子序 列,为在不失统计分布性的前提下尽可能准确地表 达小波包系数噪声分布情况,进行多次实验,最终取 子序列长度为 *l*=127,求得小波包系数序列所对应 的样本熵序列,且对其进行归一化即可得到阈值函 数调节参数序列 *s* 如图 6 所示。由图 6 可知,该小 波包系数序列中的噪声分布不均匀,因此在对低频 与中频小波包系数去噪时,调节参数较小,去噪形式 偏硬,而对其他含噪较多(调节参数较大)的系数则 去噪形式偏软。



Fig. 6 Adjust parameters of threshold function

根据以上分析利用本方法对实验信号进行去 噪,为进行直观对比,将去噪前后信号进行功率谱分 析且与其他阈值去噪方法相对比得到如图 7 所示的 结果。由于轴承外圈故障而产生周期性脉冲信号, 其表现形式为调制信号,即以轴承的转动频率为调 制频率,外圈故障频率为载波频率,形成边频带。通 常由于故障特征微弱且存在噪声干扰,轴承故障特 征及其调制特征无法清晰显露。通过信号去噪处理 后,利用功率谱分析得到轴承振动信号频谱,且由于



Fig. 7 Denoising results comparison of different algorithm

不同的去噪方法得到的功率谱分析效果不同,因此 为了直观地分析轴承振动信号去噪效果,以去噪后 信号特征频率成分的还原情况作为去噪效果评判标 准,具体分析如下。

由图 7(a)可知,原始含噪轴承振动信号功率谱 中仅能分辨出轴承基频1kHz及其倍频2kHz的 频率成分,其他有用的频率成分被噪声所淹没;通过 自相关去噪法去噪后得到的结果如图 7(b)所示,由 图可知,自相关去噪法去除了部分噪声,还原了信号 的 3 kHz 频率成分,但故障频率依然难以分辨,这 表明该信号所含噪声并非单纯的高斯白噪声,想要 得到更好的去噪结果,有必要考虑其他方法;通过 软、硬阈值函数且选用基于 Stein 无偏似然估计的 阈值确定规则(文献[8]同样采取该规则),得到去噪 结果如图 7(c)与(d)所示,硬阈值函数去噪后,能够 识别轴承故障特征频率2 552.9 Hz,但在故障频率 周围存在着无效频率成分,无法判断该故障特征的 有效性,而软阈值函数虽然滤除了大部分噪声成分, 但原始频率成分如基频1 kHz,故障特征频率 2 552.9 Hz也被扼杀严重;文献[8]阈值函数去噪结 果如图 7(e) 所示,该结果较软硬阈值函数更好,信 号的基频及其倍频 1,2,3 kHz 与轴承外圈故障频 率 $f_{\rm or} = 2552.9$ Hz 及其调制成分 3552.9 Hz 被很 好的还原,然而在3 500 Hz与高频区域仍然存在一 些无关频率成分,去噪效果有待提升;图7(f)显示 自适应阈值函数与阈值估计方法去除噪声明显且有 效地还原了原始信号的频率特征(基频及其谐波1, 2,3 kHz;故障频率及其倍频与调制成分 2 552.9, 3 552.9,5 150.8 Hz),信号失真较少,去噪效果较 好,提升了轴承故障诊断的准确性。

事实上,通过对比信号去噪前后 0.01 s 的细致 波形(如图 8 所示)也可以直观地发现,信号去噪前 其波形毛刺较多,即存在较多高频噪声干扰,而去噪 后信号细致波形毛刺较少,信号波形较为规律。

通过实验发现,本算法虽然去噪效果优于传统 去噪算法,但用时较长,计算速度较传统算法更慢, 不适用于在线实时去噪分析与故障诊断。因此,在 实际应用中,可利用传统的软、硬阈值去噪算法作为 在线实时初步分析工具,而本算法用于离线的进一 步精细分析。

4 结 论

1) 样本熵可以表征信号中不同种类的噪声含





量的大小,样本熵越大,则信号含噪越多。将样本熵 应用于小波包系数序列中,能得到其噪声分布情况, 据此对阈值函数进行自适应调整,使得其能够对含 噪较多的小波包系数进行大尺度压缩,而使含信号 较多的小波包系数得到尽可能的保护。

2)以样本熵为基准,对噪声估计 n(t)进行分析,得到最优阈值估计方法,应用于仿真信号中,证明了该阈值估计的有效性。

3)利用本方法对滚动轴承振动实验信号进行 分析,能够获得较其他方法更好的去噪效果,且有效 地还原了信号的转动特征频率与故障特征频率,提 高了轴承状态监测的准确度。

 4)本方法去噪效果较传统算法更好,但计算速 度较慢,实际应用中可将传统算法作为在线初步去
 噪算法,本方法作为离线精细算法进行配合分析。

参考文献

- [1] You Fucheng, Zhang Ying. Research of an improved wavelet threshold denoising method for transformer partial discharge signal [J]. Journal of Multimedia, 2013, 8(1):56-63.
- [2] Chen Yong, Cheng Yanan, Liu Huanlin. Application of improved wavelet adaptive threshold de-noising algorithm in FBG demodulation[J]. Optik - International Journal for Light and Electron Optics, 2017, 132 (1):243-248.
- [3] Meng Bo, Li Zhiping, Wang Haihui, et al. An improved wavelet adaptive logarithmic threshold denoising method for analysing pressure signals in a transonic

compressor[J]. Proceedings of the Institution of Mechanical Engineers Part C—Journal of Mechanical Engineering Science, 2014, 229(11): 203-210.

- [4] 刘文艺,汤宝平,蒋永华. 一种自适应小波消噪方法
 [J]. 振动、测试与诊断, 2011, 31(1):74-77.
 Liu Wenyi, Tang Baoping, Jiang Yonghua. Research on an adaptive wavelet denoising method[J]. Journal of Vibration, Measurement & Diagnosis, 2011, 31 (1):74-77. (in Chinese)
- [5] 曲巍崴,高峰. 基于噪声方差估计的小波阈值降噪研究[J]. 机械工程学报,2010,46(2):28-33.
 Qu Weiwei, Gao Feng. Study on wavelet threshold denoising algorithm based on estimation of noise variance [J]. Journal of Mechanical Engineering, 2010,46(2): 28-33. (in Chinese)
- [6] Richman J S, Moorman J R. Physiological time-series analysis using approximate entropy and sample entropy
 [J]. American Journal of Physiology Heart & Circulatory Physiology, 2000, 278(6):2039-2049.
- [7] 张建宇,张随征,管磊,等.基于多小波包样本熵的轴 承损伤程度识别方法[J].振动、测试与诊断,2015, 35(1):128-132.

Zhang Jianyu, Zhang Suizheng, Guan Lei, et al. Pattern recognition of bearing defect severity based on multiwavelet packet sample entropy method[J]. Journal of Vibration, Measurement & Diagnosis, 2015, 35 (1):128-132. (in Chinese)

- [8] Cui Huimin, Zhao Ruimei, Hou Yanli. Improved threshold denoising method based on wavelet transform[J]. Physics Procedia, 2012, 33(1):1354-1359.
- [9] Dong Wenyong, Hong Ding. Full frequency de-noising method based on wavelet decomposition and noise-type detection[J]. Neurocomputing, 2016, 214:902-909.
- [10] Zhang Bing, Xiong Jiyou, Zhang Ningsheng, et al. Improved method of processing downhole pressure data on smart wells[J]. Journal of Natural Gas Science & Engineering, 2016, 34:1115-1126.
- [11] Kumar P, Agnihotri D. Biosignal denoising via wavelet thresholds[J]. Iete Journal of Research, 2010, 56 (3):132-138.



第一作者简介:向北平,男,1974年8月 生,博士、副教授。主要研究方向为振动 测试与信号分析、机械设计及理论。曾 发表《新型梭式结构止回阀的数字仿真 与结构优化》(《四川大学学报:工程科学 版》2014年第46卷第2期)等论文。 E-mail:47058523@qq.com