• 207 •

文章编号:1671-4598(2021)11-0207-06 DOI:10.16526/j.cnki.11-4762/tp.2021.11.037 中图分类号:U270.1;TN911.72 文献标识码:A

聚类分析在消除轮轨力信号基线 漂移中的应用

农汉彪,曾巧妮

(百色学院,广西百色 533000)

摘要:针对轮轨力信号预处理过程中基点难于提取和确认的问题,根据轮轨信号幅值变化的特性,提出了基于分段数据高阶统计量聚类分析的方法来筛选出能够提取基点的数据段;轮轨力信号经过适当分段后,计算数据段的高阶统计量,按数据段的方差和峭度进行基于 OPTICS 的聚类分析,选取聚类结果靠近零点的分类所对应的数据段的中值作为基点,经曲线拟合基点后即可得到信号的基线漂移干扰;经过仿真数据和实测数据的验证和与其他现有常用方法的对比分析,结果表明该方法在均方误差和信 噪比上都优于其他方法;均方误差最高仅为其他方法的 0.47%,信噪比至少比其他方法高出 23 dB。

关键词:基线漂移,轮轨力,聚类分析,预处理,去噪

Application of Cluster Analysis in Removing Baseline Wandering of Wheel—rail Force Signal

NONG Hanbiao, ZENG Qiaoni

(Baise University, Baise 533000, China)

Abstract: Aiming at the problem that it is difficult to extract and confirm the base point in the preprocessing process of wheel rail force signal, according to the variation characteristics of wheel rail signal amplitude, a method based on cluster analysis of higher—order statistics of segmented data is proposed to screen the data segments that can extract the base point. After the signal is properly segmented, the high—order statistics of the data segment are calculated, and the clustering analysis based on OPTICS is carried out according to the variance and kurtosis of the data segment. The median value of the data segment that corresponding to the classification of the clustering result which close to the zero point as the base point, and the baseline wandering interference of the signal can be obtained after curve fitting the base point. The results of verification with simulation data and measured data and comparative analysis with other existing common methods, show that this method is superior to other methods in mean square error and signal—to—noise ratio. The maximum mean square error is only 0. 47 % of that of other methods, and the signal—to—noise ratio is at least 23 dB higher than that of other methods.

Keywords: baseline wandering; wheel-rail force; cluster analysis; preprocessing; denoising

0 引言

轮轨相互作用力监测是轨道交通车辆运行安全状态中 非常重要的方法和途径。通过轮轨力的监测可以实时了解 车辆运行过程中的减载率,脱轨系数等。在基于地面的轮 轨力检测中,通常通过测量钢轨的动态响应来测量车轮通 过时的轮轨相互作用力^[1]。而在现场实测时,由于钢轨支 撑条件会存在一定的不确定性,比如悬枕,扣件松弛,以 及地基弹性突变等情况,使得轮轨力测试信号中会出现基 线漂移的干扰^[2],基线漂移的出现影响了后续的轮轨力信 号的定量分析以及对列车运行状态的安全评估,所以必须 予以去除。

去除基线漂移方法在心电图信号处理^[3-4],脉搏信号处 理^[5],眼电图处理^[6]等生物医学领域中广泛应用,在光学测 试^[7]和地震监测^[8]中也有了相关的文献论述。而针对轮轨力 信号降噪和基线漂移的消除,目前国内外较少涉及。

文献 [9] 中提出了基于分段数据中值聚类的算法,该 算法将信号数据分成很多分段,取分段数据的中值,将中 值的时间间距与幅值间距作为距离测度进行最近距离聚类, 得到了能代表基线漂移的基点,拟合基点得到基线漂移干 扰。算法效果良好,然而算法中需要设置不同测度之间的 协调因子,且该因子的设置与数据长度,采样频率,数据 幅值等因数相关。文献 [10] 利用 LabVIEW 提供的高级信 号处理工具包和数字滤波器设计工具包对实测轮轨力应变 信号进行处理,内容包括消除基准漂移、去除宽带噪声和 提取信号特征等。文献 [11] 提出一种小波包降噪算法与 经验模态分解 (EMD) 相关算法相结合的数据处理方法, 有效地消除轮轨力在采集过程中随机白噪声的干扰和抑制

收稿日期:2021-08-12; 修回日期:2021-09-23。

作者简介:农汉彪(1982-),男,广西田阳人,博士,工程师,主要从事轨道交通车辆试验测试技术方向的研究。 引用格式:农汉彪,曾巧妮.聚类分析在消除轮轨力信号基线漂移中的应用[J].计算机测量与控制,2021,29(11):207-212. 模态的混叠。文献 [12] 提出了基于集合经验模态分解 (EEMD) 与小波变换相结合的去噪方法。对含有基线漂移 的分量通过小波变换进行分解,将代表基线漂移的趋势项 置零达到去除基线漂移的目的。

近年来,基线漂移的去除方法趋向于将原信号通过不同的算法分解成纯净信号部分和干扰成分,主要算法有基 EMD 及其改进算法,如 EEMD^[13],CEEMDAN^[14]等,有经 验小波分解(EWT)^[15],傅里叶分解^[16],变分模态分解 (VMD)^[17]等,而信号的分解总会存在频率混叠问题,传统 的方法有 FIR 或 IIR 的方法,中值滤波^[18],曲线拟合^[19]和 数学形态学滤波^[20]等。曲线拟合方法是将信号中能代表基 线漂移走势的基点拟合成曲线并从原始信号中提取和确定 基点比较困难。本文根据轮轨力信号短时分段数据的高阶 统计量分布和基线漂移的关系,提出了基于分段数据高阶 统计量聚类分析的方法获取基点所在的数据分段,以数据 分段中值作为基点,再拟合即可得到基线漂移。

1 算法思想与实现

钢轨的模型通常采用连续弹性基础梁进行表示,钢轨 垂直位移与移动载荷的大小,速度,位置之间关系表 示为^[21]:

$$EI \frac{\partial^4 y(x,t)}{\partial x^4} + m \frac{\partial^2 y(x,t)}{\partial t^2} + ky(x,t) = P\delta(x-ut)$$
(1)

其中: *EI* 为钢轨的抗弯刚度, *k* 为地基的弹性系数, *m* 为单位长度钢轨质量, *P* 为钢轨上载荷的大小, *x* 为载荷作 用点与观测点之间的距离, *v* 为载荷移动的速度, δ(•) 为狄拉克函数。

静态或准静态时,在多个载荷的同时作用下,钢轨上 某观测点的剪力为:

$$Q(t) = -\frac{1}{2} \sum_{i=1}^{n} P_i(t) \exp(-\frac{x_i(t)}{L}) \cos\frac{x_i(t)}{L}$$
(2)

其中: $L = \sqrt[4]{\frac{4EI}{k}} P_i$ 为第 i 载荷的大小, x_i 为第 i 载荷作用点与观测点之间的距离。

在实测信号中,测试系统采集到信号除了钢轨载荷引 起的响应之外,还包含了基线漂移成分和其他的随机干扰, 即采集到的信号 *f*(*t*)可以表示为:

f(t) = Q(t) + BW(t) + n(t) (3)

其中: Q(t) 为车轮载荷所引起的剪力测量值; BW(t) 为 基线漂移信号, n(t) 为白噪声干扰信号,当所有 $|x_i| > 3\pi L/2$ 时, Q(t) 降为峰值的 1%以内,此时 $BW(t) + n(t) \ge Q(t)$,称无载荷区间即:

$$f(t) \approx BW(t) + n(t) \tag{4}$$

根据随机噪声的对称性和基线漂移信号的缓变性,当 所有 $|x_i| > 3\pi L/2$ 时,短时间段内测试信号的中值即可表 示为基线漂移成分的幅值,称为基点。而当存在 $|x_i| <$ 3πL/2时,测量值 f(t) 包含钢轨对载荷的响应,其数据段 的中值不能表示为基点。拟合所有基点即可达到信号的基 线漂移成分。而前提是正确筛选出能被视为基点的数据段 中值数据。

对于 f(t), 由于 Q(t), BW(t), n(t) 相互独立, 则有 f(t) 的方差:

$$D(f(t)) = D(Q(t) + BW(t) + n(t)) =$$

$$D(Q(t)) + D(BW(t)) + D(n(t))$$
(5)

BW(t)为缓变信号,在短时间内 BW(t)为可视为常数,即D(BW(t)) = 0,而白噪声的方差D(n(t))可视为常数C;从而有当所有 $|x_i| > 3\pi L/2$ 时:

$$D(f(t)) = 0 + 0 + C = C$$
(6)

而在 | x_i | <3 $\pi L/2$ 时:

$$D(f(t)) = D(Q(t)) + 0 + C = D(Q(t)) + C$$
(7)

另外,同样可以看到,在短时间内信号的峭度值:

$$K(f(t)) = \frac{E(f(t) - \mu)^4}{\sigma^4}$$
(8)

其中 μ 为f(t)的数学期望, σ 为f(t)的标准差,由于 Q(t), BW(t), n(t)相互独立,有:

K(f(t)) = K(Q(t)) + K(BW(t)) + K(n(t)) (9) 短时间内 BW(t) 是均匀分布,即 K(BW(t)) = 0,n(t) 为正态分布, K(n(t)) = 3; 从而有当所有 | x_i | >3 πL /

$$K(f(t)) = 0 + 0 + 3 = 3 \tag{10}$$

而在 | x_i | $< 3\pi L/2$ 时:

2时:

 $K(f(t)) = K(Q(t)) + 0 + 3 = K(Q(t)) + 3 \quad (11)$

联合方差与峭度,无载荷期间内数据段的 D(f(t)), 和 K(f(t))将主要分布在以方差和峭度为坐标轴的点(0, 3)和点(C,3)附近,而承载期间的坐标位置将远离这两 点坐标。根据以上分析,可以通过对分段数据的方差和峭 度进行基于密度的聚类分析来筛选出无载荷期间的数据段。

基于密度的聚类分析可以将具有相同或相近特征的对 象样本之间的分为同一类,即在某类别任意样本周围一定 空间范围内一定有同类别的样本存在,能有效区分有载荷 期间数据段和无载荷期间数据段。基于密度的聚类分析的 常用算法有 DBSCAN,OPTICS 和 DENCLUE 等。DB-SCAN 需要设定点数阈值 minPts 和邻域半径,全局密度, DENCLUE 需要设定密度阀值和参数,这两种算法的效果 对参数非常敏感,需要确定合适的参数才能得到理想的效 果。OPTICS 是 DBSCAN 的一种改进方案,也要设定点数 阈值 minPts 和邻域半径,但聚类结果对参数不敏感。

基于以上分析,应用基于密度的聚类分析可以有效找 到信号中基点,从而实现基线漂移的消除。该方法的具体 过程和步骤如下:

1) 对测量数据进行分段。前后数据段之间可以部分重
 叠,以加大数据段数量和保证数据的连续性;

2) 计算每数据段方差和峭度值,并分别进行归一化

(14)

处理;

3)针对归一化方差和峭度两个维度进行基于密度的聚
 类分析;

 4)筛选出聚类结果中处于零点附近的一个分类或几个 分类;

- 5) 对选用分类所对应的数据段取中值,即为基点;
- 6)曲线拟合所有基点得到基线漂移干扰。

分段时,数据段的大小需要根据信号自身的成分来确 定。涉及到信号数据的采样频率,通过列车的转向架间距, 轴距和运行速度等。分块大小的最优设置是能将信号的负 荷状态和非负荷状态完全分开,这是理想的情况,而实际 上由于信号数据初始采集相位未知且列车运行速度并不恒 定原因,无法在未识别压头时刻的情况下实现理想分段。 而总会出现部分分段横跨负荷状态和非负荷状态两种状态 的情况。根据数据试验结果一般采用的分大小可为:

$$N = \frac{L_{\rm B-B} - L_{\rm A-A}}{v} \times Fs \tag{12}$$

其中: L_{B-B}为列车车辆定距(转向架中心距), L_{A-A}为 转向架轴距, v 为列车通过的平均速度, Fs 为信号的采样频 率。在数据量较少的情况下,可以通过使用较小的步进大 小来获取更多的数据分段。分段数据的大小和步进长度也 可以通过信号峰值检测的结果来自动设置, 假设列车通过 测量点的速度恒定的情况, 分段大小的设定可设为一个波 峰宽的数据长度。

分段数据的方差和峭度都是非负数值,但具有不同的 数值范围,基于密度的聚类需要考虑数据点之间的距离测 度或相似度测度,为了均衡方差和峭度在距离测度上的贡 献,需要对方差和峭度进行归一化处理。归一化时可以将 最大方差和最大峭度置为1,最小值置为零,其他进行线性 化处理。

在聚类分析时,选用对参数不敏感的 OPTICS 算法, OPTICS 所需要点数阈值 minPts 和邻域半径 ε 两个参数可 以通过自适应的方式来确定。邻域半径 ε 可以取所有数据样 点两两间距测度密度分布函数的最快下降梯度点所对应的 距离测度,minPts 可以取使得当前 ε 下所有分类间距总和 最小的阈值。

2 仿真数据分析

为了定量分析基线消除方法的效果,预先准备一组零 漂移轮轨垂直力信号和一组已知基线漂移信号。假设列车 为两轴转向架,车辆定距为7.67 m,转向架轴距为1.75 m, 车辆长度为 11.986 m,列车低速经过观测点的车速为 4 km/h,信号采样频率为1000 Hz,数据长度为 60 000 点, 各轮轨力信号峰值随机分布在 1.4~1.7 之间零漂移轮轨垂 直力信号图 1 所示,数据本身具有很低的频率成分,该频 率成分与列车转向架间距,轴距与列车速度都有关系。基 线漂移信号为:

$$BW(t) = 0.15 * (\sin(0.15\pi t) + \sin(0.09\pi t + 0.2\pi))$$

采用与力信号相同的采样间隔。基线及与力信号叠加 后的结果如图 2 所示。



图 2 基线及与力信号叠加的结果

常用于定量分析滤波算法性能评估参数主要有均方误差(MSE)和信噪比(SNR)等^[22]。对于性能评价,MSE 值越小越好,SNR 越大越好。它们的定义分别为:

$$MSE = \frac{\sum_{n=1}^{N} (x(n) - \hat{x}(n))^2}{N}$$
(15)

$$SNR = 10\log_{10}\left(\frac{\sum_{n=1}^{N} \hat{x}^{2}(n)}{\sum_{n=1}^{N} (x(n) - \hat{x}(n))^{2}}\right)$$
(16)

其中:x(n)为原始信号, $\hat{x}(n)$ 为处理后的信号。

根据分段大小的分析,仿真数据可以采用分段大小为 500数据点对含有基线漂移的信号数据进行分段,分段窗口 步进长度为 250。分段后的数据分段中值如图 3 所示,大部 分分段数据的中值和已知的基线漂移重合,另一部分偏离 较远。

随后对所有数据分段统计数据分段的方差和峭度,得 到的方差和峭度数据如图4所示。

方差数值在 0~1 范围之内,而峭度数值在 0~80 范围 之内,分别对方差和峭度进行线性归一化,然后对归一化 后的数据进行聚类分析得到如图 5 的分类结果。

图 5 中,数据点被分成了两个分类和部分未分类样点。 根据轮轨力数据的特点,选用分类1 作为无载荷期间的数



据段。对分类1所对应的数据分段中值即为信号数据的基 点。基点数据与目标基线漂移的位置关系如图6所示。

为了对比分析,利用移动中值滤波,形态学滤波, (BEADS, baseline estimation and denoising with sparsity)^[23], CEEMDAN, EWT, VMD,零相位滤波器以及本 文提出的方法分别对图 2 中含有基线漂移的力信号进行处 理。在这些方法可分为两类,一类是 CEEMDAN, EWT,



图 6 选用基点与原始信号

VMD 和零相位滤波器,另一类是移动中值滤波,形态学滤 波,BEADS 和本文的方法。第一类主要是信号分解与重构 的方法和数字滤波的方法,这些方法会将所有的测量数据 值进行计算,这使得在有载荷期间的基线漂移的提取会受 到载荷大小的影响。各种方法通过优化设置参数得到最优 的结果如图 7 所示。所得结果与目标基线存在较大的误差, 不适合后续的定量分析。





第二类方法中,设置移动中值滤波的移动窗口大小为 700数据点,移动步进为一个数据点,然后再经过1Hz低 通滤波器得到最优的平滑基线;数学形态学滤波采用长度 为350数据点的横条型结构分别进行开闭运算然后取两部 分和的一半,再经过1Hz低通滤波器得到最优的平滑基 线;依据文献[23]BEADS方法采用截止频率为0.8Hz, 阶数为1,不对称参数为12,正则化参数为0.01。由于 BEADS方法对任何信号提取的基线首尾均为零,需要先对 原信号的首尾置零提取趋势线再进行提取,才能得到最佳 的评估效果。不同方法提取的基线漂移与目标基线的误差 如图8所示

从基线漂移误差曲线看到,移动中值滤波和形态学滤 波具有形同的变化规律,即会在观测点负载时会有较大误 差而其他时刻的误差趋于零,BEADS方法在信号的两头出 现较大的误差,而中间部分误差较小。而本文的方法提取 基线的误差相对其他3种方法都要小。原始带基线漂移的



图 8 提取的基线漂移与目标基线的误差曲线

信号在经过不同方法去除基线前后的相似测度参数如表 1 所示。

表 1	不同方法	去除基线	漂移的效果
1. 1		A 120 45 -73	125 12 11 26 25

去除方法	SNR/dB	MSE
移动中值	33.861 39	5.66786E-05
形态滤波	34.695 43	4.67751E-05
BEADS	48.778 86	1.82672E-06
本文方法	72.036 53	8.62791E-09

从表1的数据结果表明,本文的方法对于仿真数据的 消除基线漂移的两种评价指标均优于形态滤波等其他3种 方法。本文方法的均方误差 MSE 仅约为次好的 BEADS 算 法的 0.47%,而信噪比 SNR 则高出 23 dB。

3 实测数据分析

实测数据中,除了基线漂移的干扰之外还会有其他噪声的存在,如工频干扰和随机干扰等。如图9所示为一组数据现场实测轮轨力信号。



依据算法,首先获取数据峰值宽度以确定数据分段的 大小,接着对实测信号进行数据分段,统计分段数据的方 差和峭度,基于方差和峭度统计数据段之间距离,按距离 分布函数获取最优聚类领域半径,然后进行基于 OPTICS 算法进行聚类分析,得到聚类结果如图 10 所示,选用分类 1 所对应的数据段,以该分类数据段的中值作为基点,拟合





图 11 实测轮轨信号基线漂移的基点的选取

实测信号没有预知的基线漂移或纯净的信号数据,无 法进行量化的评估。为了验证本文方法的效果,采用移动 中值滤波,形态学滤波,BEADS方法分别对实测信号进行 处理,对比的方法通过调整相应的参数使得处理结果达到 最佳效果。图12为不同方法消除实测轮轨垂直力信号基线 漂移的效果。其中最上方的曲线为实测的原始信号,往下 分别为移动中值滤波加低通滤波的方法,形态学滤波加低 通滤波的方法,BEADS方法以及本文方法去除基线漂移后 的信号波形图。

从图上看,本文的方法和 BEADS 方法于实测数据中的 基线漂移的消除效果明显优于前两种方法。本文的方法和 BEADS 方法都能有效地去除基线漂移的干扰,仅从视觉上 看,两者无法区分两者优劣情况。从计算量上,由于本文 的方法采用分段的方法,使得整个计算过程的计算量相比 与移动中值滤波、形态学滤波和 BEADS 方法大大的减小, 提高了计算效率。

4 结束语

本文针对轮轨垂向力信号的基线漂移消除问题,提出 了基于分段数据高阶统计量聚类分析的方法来确定信号基 点所在的数据段,统计选用数据段中值作为基点,曲线拟 合基点即可得到信号的基线漂移干扰。文中通过仿真数据



和现场实测信号从定量和定性两种方式对所提出的方法进 行了验证,并与现有常用的基线漂移去除方法进行了横向 比较。验证结果都表明相比于其他方法,本文的方法在定 量上具有更高的信噪比和更小的均方误差,在对实测信号 的定性分析上同样具有很好的效果,而且计算量更小,速 度更快。本文方法的参数设置可以根据信号的特性自适应 设置,不需要过多的人为参与。该方法同样适用于其他包 含有基线偏移的测试信号的处理,比如心电信号,光谱信 号,脉搏信号等。

参考文献:

- [1]朱 涛,肖守讷,阳光武,等. 车辆垂向轮轨力识别方法与试验[J]. 振动.测试与诊断,2015,35 (5):892-896.
- [2]李 洲,郑树彬,柴晓冬.基于降维状态观测器的轮轨力估计方法研究[J].计算机测量与控制,2014,22 (9):2893-2895.
- [3] WANG X, ZHOU Y, SHU M, et al. ECG Baseline Wander Correction and Denoising Based on Sparsity [J]. IEEE Access, 2019, 7: 31573 - 31585.
- [4] ROMERO F P, PI? OL D C, VáZQUEZ-SEISDEDOS C R. DeepFilter: An ECG baseline wander removal filter using deep learning techniques [J]. Biomedical Signal Processing and Control, 2021, 70 (102992): 1-10.
- [5] 苏志刚,吕江波,郝敬堂,等.改进的经验模态分解法去除脉 搏波基线漂移 [J]. 计算机测量与控制,2018,26(7):229 -232.
- [6] NATHANIEL B, TRACEY A C, KENNETH P C. A comparison of EOG baseline drift mitigation techniques [J]. Biomedical Signal Processing and Control, 2020, 57 (101738): 1-9.
- [7] 李 玲. 近红外光谱基线漂移校正的方法研究 [J]. 激光杂志,

2019, 40 (2): 153-156.

- [8] 吴先敏,牛超,卞 晶,等.地震波基线漂移校正及结构地 震响应分析 [J].水资源与水工程学报,2019,30 (2):186-190.
- [9] 农汉彪. 轮轨垂向载荷连续测量与识别方法研究 [D]. 成都: 西南交通大学, 2012.
- [10] 赖建飞, 雷晓燕, 刘庆杰. LabVIEW 小波变换在轮轨力应变 信号处理中的应用 [J]. 城市轨道交通研究, 2015, 18 (2): 1-6.
- [11] 张绪景, 雷晓燕, 刘庆杰. 基于小波包与改进 EMD 的轮轨力 信号降噪 [J]. 噪声与振动控制, 2016, 36 (5): 104-107.
- [12] 刘庆杰,黄 辉,雷晓燕.基于集合经验模态分解和小波变 换的轮轨力应变信号降噪 [J].城市轨道交通研究,2016,19 (11):26-29.
- [13] MARIYAPPA N, SENGOTTUVEL S, PARASAKTHI C, et al. Baseline drift removal and denoising of MCG data using EE-MD: Role of noise amplitude and the thresholding effect [J]. Medical engineering & physics, 2014, 36 (10): 1266-1276.
- [14] YAO L, PAN Z. A new method based CEEMDAN for removal of baseline wander and powerline interference in ECG signals [J]. Optik, 2020, 223 (165566): 1-13.
- [15] BODA S, MAHADEVAPPA M, DUTTA P K. A hybrid method for removal of power line interference and baseline wander in ECG signals using EMD and EWT [J]. Biomedical Signal Processing and Control, 2021, 67 (102466): 1-34.
- [16] SINGHAL A, SINGH P, FATIMAH B, et al. An efficient removal of power—line interference and baseline wander from ECG signals by employing Fourier decomposition technique [J]. Biomedical Signal Processing and Control, 2020, 57 (101741): 1-9.
- [17] DRAGOMIRETSKIY K, ZOSSO D. Variational mode decomposition [J]. IEEE Transactions on Signal Processing, 2014, 62 (3): 531-544.
- [18] 蔡 坤,陆尧胜.基于中值滤波的心电基线校正方法的研究 [J]. 医疗设备信息,2004,19(2):5-7.
- [19] 万相奎,唐文普,张 赖,等.改进的三次样条插值心电基 线漂移滤波法 [J]. 生物医学工程学杂志,2016,33 (2): 227-231.
- [20] YU P, LU D, JIN Z L, et al. Removal Method of Baseline Drift from ECG Signals Based on Morphology Filter [J]. Applied Mechanics and Materials, 2013, 429: 1691-1695.
- [21] 翟婉明. 车辆一轨道耦合动力学(上册)[M]. 北京:科学出版社,2015.
- [22] 陈 苹,叶继伦,张 旭,等.心电信号中基线漂移的去除 方法比较及算法实现 [J].中国医疗器械杂志,2018,42 (5):326-329.
- [23] NING X, SELESNICK I W, DUVAL L. Chromatogram baseline estimation and denoising using sparsity (BEADS) [J]. Chemometrics and Intelligent Laboratory Systems, 2014, 139: 156-167.