

智能手机检测车辆振动加速度数据预处理方法

从建力^{1,2} 王 源^{1,2} 杨翠平^{1,2} 王 平^{1,2} 李成辉^{1,2}

(1. 西南交通大学土木工程学院, 成都, 610031; 2. 西南交通大学高速铁路线路工程教育部重点实验室, 成都, 610031)

摘 要: 通过所开发的手机端软件从智能手机中提取车辆振动加速度的过程中, 为正确评价轨道平顺状态及车辆运行舒适性, 需保障车辆振动加速度数据质量。本文依据概率统计方法与小波分析方法建立了异常值识别模型, 采取中值滤波和小波滤波剔除因手机传感器性能稳定性和测试环境变化引起的随机误差, 并结合成都地铁手机检测数据, 验证两种滤波方法对手机检测数据随机误差的处理效果。实例分析表明: 基于异常值识别模型可准确提取手机检测数据异常值位置, 借用中值滤波和小波滤波剔除因外界环境变化引起的随机误差, 保证了手机检测数据真实反映车体振动响应, 为正确评价轨道平顺状态和车辆运行舒适性提供数据支撑。

关键词: 智能手机; 车体振动响应; 异常值; 数据预处理

中图分类号: U212.34 **文献标志码:** A

Data Preprocessing Method of Vehicle Vibration Acceleration by Smartphone

Cong Jianli^{1,2}, Wang Yuan^{1,2}, Yang Cuiping^{1,2}, Wang Ping^{1,2}, Li Chenghui^{1,2}

(1. School of Civil Engineering, Southwest Jiaotong University, Chengdu, 610031, China; 2. Key Laboratory of the Ministry of Education, High Speed Railway Engineering, Southwest Jiaotong University, Chengdu, 610031, China)

Abstract: When extracting vehicle vibration acceleration from smart phone through software developed, the data quality of the vibration acceleration of the vehicle should be guaranteed in order to correctly evaluate the track smoothness and vehicle running comfort. This paper establishes an abnormal value recognition model based on the methods of probability-statistics and wavelet, in which median filter and wavelet filter are used to eliminate the random errors caused by the performance stability of mobile sensors and the change of testing environment. The effect of two filtering methods on the random error of mobile phone detection data is verified by combining with the detection data from Chengdu metro. The analysis result shows that the abnormal value location of mobile phone detection data can be accurately extracted based on the abnormal value recognition model. The mobile phone detection data truly reflect the vibration response of the car body by eliminating random errors caused by external environmental changes using methods of median filter and wavelet filter, so the model can be used to correctly evaluate the track smooth state and vehicle operating comfort.

Key words: smart phone; vibration response of car body; abnormal value; data preprocessing

引言

基于列车车体振动加速度来评估轨道平顺状态和车辆运行舒适性^[1],对铁路线路养护维修来说是一种便捷、有效的方法,但这类评估方法对车体加速度数据的体量需求巨大。为了适应城市轨道交通、高速铁路的轨道平顺性评估,本文考虑借助智能手机作为感知终端^[2],通过铁路职工的智能手机所测量得到的加速度数据,为铁路线路养护提供重要的数据源。相对于精密加速度传感器和添乘仪,使用手机测试车体振动响应的主要优点是不会影响行车安全和旅客乘车体验,并且成本低、数据体量巨大。

该项工作面临的主要问题是由于测试环境的复杂性、测量终端的流动性等原因导致智能手机所获取的数据存在较强的误差干扰,原因包括:(1)因手机后台运行多种 App 导致手机采样频率不稳定;(2)因乘客上下车和随意走动引起的偶然干扰;(3)因手机内置传感器信噪比低引起的随机误差。因此必须经过系统性的数据预处理才可发挥该类数据的价值。本文的主要工作即是针对智能手机测量数据,研究其数据清洗方法,以增强该类数据的实用价值。

合理剔除测量数据异常值是各工程领域遇到的问题,对于不能确定总体分布类型的数据,需做适当的数据变换使数据正态化^[3]。为发挥数据价值,需对数据进行清洗^[4-8],其过程一般包括两个方面:(1)异常值的识别定位,如设定阈值筛选异常值^[9];(2)异常值的处理,如小波分析具有较好的信号降噪效果^[10-11],中值滤波可剔除信号中较大误差^[12-13]。智能手机传感数据可能的异常类型较多,而上述文献中采用单一的异常值识别和处理方法,无法直接满足手机检测数据清洗需求,需要研究系统性的数据异常识别-分类处理框架。

基于此,本文基于统计方法,根据不同的异常值产生原因,首先提出数据异常值识别方法:P值筛选、一阶变化率识别与小波分析3种方法相结合;进而提出对应的数据异常处理方法:区段删除、中值滤波、小波滤波3种处理方法相结合。实例分析表明,本文方法处理后的智能手机传感数据具有较稳定的时域、频域特性,可为评估轨道平顺状态和列车运行品质提供可靠的数据保障。

1 数据预处理模型

基于智能终端采集地铁车辆振动加速度时,为使数据能够服务于地铁系统,包括评价轨道平顺状态、车辆舒适性^[14-16],需对手机检测数据进行质量分析。智能手机主要承担数据采集任务,添乘人员将检测数据及时上传至服务器,依靠服务器较大的存储空间和较快的计算速度进行数据预处理工作,具体内容包括:(1)因手机采样频率不稳定,需基于检测数据进行重采样得到等时间间隔数据序列;(2)受测试环境变化及传感器信噪比低等多种因素影响,手机检测数据含有随机误差,需对数据异常值进行识别与分类;(3)针对数据异常值进行分类处理;(4)数据清洗后应用分析。本文主要阐述前3部分,数据预处理流程如图1所示。

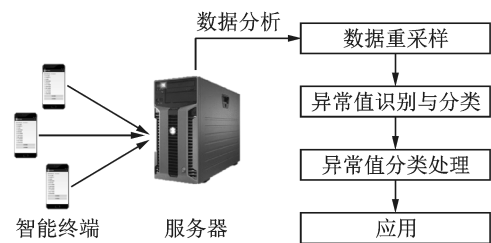


图1 数据预处理流程

Fig.1 Data preprocessing process

2 数据重采样

因手机后台同时运行多种应用软件引起手机采样频率不稳定,其实际采样频率 f 在理论采样频率 F 上下变动,造成手机检测数据并不是等间隔时间序列。本文采用线性插值法对手机检测原生数据进行重新采样,生成等间隔时间序列,采样原理示意图如图2所示。

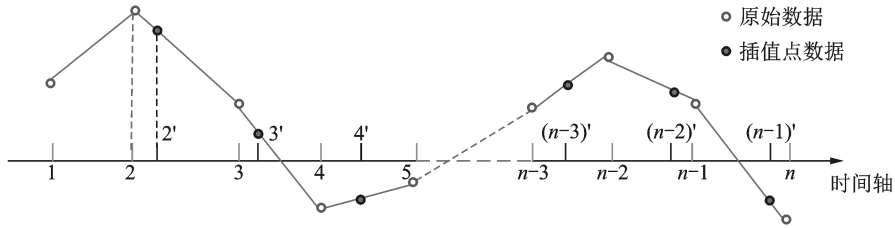


图2 线性插值示意图

Fig.2 Linear interpolation diagrammatic sketch

等间隔时间序列计算公式如下

$$y_i = x_i + \frac{x_{i+1} - x_i}{t_{i+1} - t_i} (t_i - t_i) \tag{1}$$

式中: y_i 为原生数据*i*位置插值后数值; x_i, x_{i+1} 为原生数据*i, i+1*位置数值; t_i, t_{i+1} 为原生数据*i, i+1*位置数据采集时刻; t_i 为理论采样频率*F*下*i*位置处数据采集时刻。

3 异常值识别与分类

因手机检测数据中的异常值是由多种原因引起的,本文基于*P*值筛选识别出的数据异常值定义为第一类异常值;基于检测数据一阶变化率并结合“ 3σ 准则”识别超出设定阈值的数据定义为第二类异常值;基于小波分析分解因传感器信噪比低引起的随机误差定义为第三类异常值。上述异常值识别方法描述如下:

(1)*P*值筛选:设置数据始端至末端的移动窗口长度为*N*,通过设定阈值*T*并计算移动窗口中超出阈值*T*范围的概率*P*,其公式如下

$$P = \frac{n(x_i > T)}{N} \tag{2}$$

$$x_i = \begin{cases} 1 & P > P_0 \\ 0 & P < P_0 \end{cases} \tag{3}$$

式中:*n*为窗口元素*x_i*超出阈值*T*的个数;*N*为移动窗口长度;*P₀*为自定义概率常数。*x_i*为1时表示手机检测数据属于第一类异常值。

(2)一阶变化率识别:根据手机检测数据一阶变化率并结合“ 3σ 准则”将超出阈值的一阶变化率视为异常值。其一阶变化率计算公式如下

$$y_i = \frac{x_{i+1} - x_i}{t_{i+1} - t_i} \tag{4}$$

式中: x_i, x_{i+1} 为*i, i+1*位置的手机检测数据; t_i, t_{i+1} 为相应手机检测数据时间; y_i 为*i*位置处一阶变化率值。

根据“ 3σ 准则”设置一阶变化率超限阈值,其阈值范围可表示为

$$\mu - 3\sigma < y_i < \mu + 3\sigma \tag{5}$$

$$\mu = \frac{1}{N_1} \sum_{i=1}^N y_i \tag{6}$$

$$\sigma = \sqrt{\frac{1}{N_1} \sum_{i=1}^N (y_i - \mu)^2} \tag{7}$$

式中: μ 为一阶变化率均值; σ 为一阶变化率标准差; y_i 为 i 位置一阶变化率值; N_1 为数据一阶变化率长度。

由于手机检测数据受测试环境影响,手机检测数据局部会发生突变,表现为一阶变化率幅值超限。为了避免将幅值较大的手机检测有效数据识别为异常值,当手机检测数据一阶变化率满足式(5)时,可保证手机检测数据具有较高可信度(99.74%)。由式(8)识别一阶变化率超限位置。

$$y_i = \begin{cases} 1 & y_i < \mu - 3\sigma \text{ 或 } y_i > \mu + 3\sigma \\ 0 & \mu - 3\sigma < y_i < \mu + 3\sigma \end{cases} \quad (8)$$

式中 y_i 为 1 时表示 i 位置处手机检测数据为第二类异常值。

(3)因手机加速度传感器信噪比低引起的随机误差,该类随机误差识别方法如下:根据手机采样频率 F 确定小波分解层数 n ,将手机检测信号分解为低频成分和细节成分,这里的细节成分包含因手机加速度传感器信噪比低引起的成分。由于在时域分析中离散傅里叶变换没有局部化能力,而小波分析可对时域信号局部信息进行分解,这里借鉴小波分析具有的局部化能力,对手机检测数据局部波动较大位置进行多层分解。小波尺度 a 和平移 t 是基小波两个重要参数,这里选择经典小波尺度 $a=2$, t 为手机采样间隔 Δt ,便可得出小波分解过程中每层对应的中心频率 f_c ,其计算公式如下

$$f_c = \frac{F}{2^n} \quad n = 1, \dots, i, \dots, N \quad (9)$$

式中: F 为手机标准采样频率; n 为小波分解层数。

手机检测数据可视为离散数据信号,且信号能量有限,即

$$W_x = \sum_{i=1}^N |x_i|^2 \Delta t < +\infty \quad x(t) \in L^2(R) \quad (10)$$

式中: $L^2(R)$ 为 R 上平方可积函数构成的函数空间; W_x 为信号能量且有限。

这里选择基小波 $\psi_{a,b}(t)$ ($a=2, b=\Delta t$),将基小波经过伸缩和平移后,就可以得到一个小波序列,即

$$\psi_{2,\Delta t}(t) = \frac{1}{\sqrt{2}} \psi\left(\frac{t-\Delta t}{2}\right) \quad (11)$$

将基小波 $\psi_{2,\Delta t}(t)$ 与手机检测数据进行卷积实现分解

$$\langle x_i, \psi_{2,\Delta t} \rangle = \frac{1}{\sqrt{2}} \sum_{i=1}^N x_i \psi\left(\frac{t-\Delta t}{2}\right) \Delta t \quad (12)$$

式中: x_i 为长度 N 的离散随机信号; $\langle x_i, \psi_{2,\Delta t} \rangle$ 为 x_i 与 $\psi_{2,\Delta t}$ 的内积。

4 异常值处理

本文根据3种类型异常值的分布特征,确定了相对应的异常值处理方法,具体如下:

(1)第一类异常值幅值波动较大,波动个数与窗口元素个数比例 P 超出设定阈值 P_0 ,且幅值远远超出车体正常振动加速度范围,这是受明显干扰因素影响的数据,不能真实反映车体振动响应,采取区段删除方法直接剔除。

(2)第二类异常值幅值波动较小,该类异常值含有车体振动响应有效成份,可采取中值滤波方法滤除该类型异常值对车体振动响应的影响,其原理描述如下:该方法需提前设置长度为 N (奇数)的移动窗口,以异常值所在位置为窗口中心位置,将窗口内手机检测数据按升序或降序重新排列,用产生的新序列数据的中值元素代替异常值,可表示为

$$x_i = \text{median}([x_{i-(N-1)/2}, \dots, x_i, \dots, x_{i+(N-1)/2}]) \quad (13)$$

式中: x_i 为移动窗口内 i 位置对应的元素; $x_{i-(N-1)/2}, x_{i+(N-1)/2}$ 为移动窗口两端元素; median 表示取移动窗口中值元素; x_i 为窗口元素重新排序后的中值元素。

(3) 本文采用小波滤波去除因手机加速度传感器信噪比低引起的随机误差, 其步骤为: 首先选择合适的小波函数进行分解计算; 接着选择半软阈值函数对各个分解尺度下的细节成份进行阈值量化处理; 最后将小波分解的最低层低频系数和阈值量化处理后的各层细节成份进行重构, 其基本原理数学描述如下。

半软阈值函数

$$\eta(w) = \text{sgn}(w) \frac{T_2(|w|) - T_1}{T_2 - T_1} I(T_1 < |w| < T_2) + w I(|w| > T_2) \quad (14)$$

式中: sgn 为阶跃函数; w 为高频系数; I 为两数取小; T_1, T_2 为常数, 且 $0 < T_1 < T_2$ 。当 $w > 0$, $\text{sgn} = 1$; $w = 0$, $\text{sgn} = 0$; $w < 0$, $\text{sgn} = -1$ 。

由小波重构基本关系式 $V_{n-1} = V_n \oplus W_n$ 可知

$$x^{n-1} = \sum_i x_i^n \phi_{mi}(t) + \sum_i d_i^n \psi_{mi}(t) \quad (15)$$

式中: x^{n-1} 为第 $n-1$ 层待重构信号; x_i^n 为已完成小波重构的第 n 层低频系数; d_i^n 为已完成小波重构的第 n 层高频系数; $\phi_{mi}(t)$ 为尺度函数; $\psi_{mi}(t)$ 为小波基函数。

5 实例分析

5.1 现场实测数据

本试验测试路线为成都地铁1号线火车北站至高新站, 共12个车站, 11个区段, 将智能手机放置在相对稳定的车厢地板上采集车体振动响应, 依据文献[14, 16]确定手机安放位置, 手机标准采样频率为100 Hz, 手机检测数据如图3(a)所示。

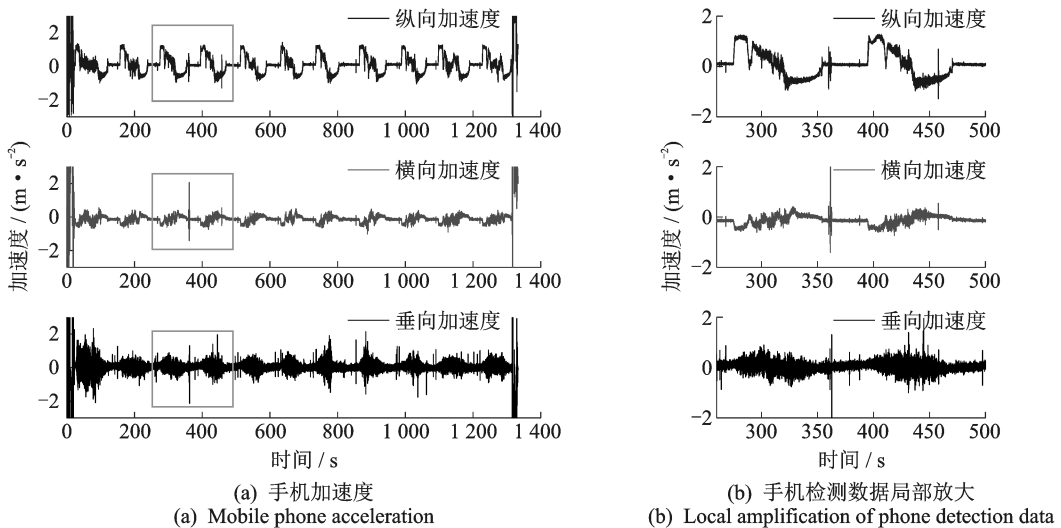


图3 手机检测原生数据

Fig.3 Mobile phone detection of native data

结合现场测试情况可知,因手机测试开始和结束时,测试人员触碰手机点击“启动测试”和“关闭测试”,数据始端和末端出现较大波动;在测试过程中受乘客上下车和随意走动影响,手机检测数据局部出现较多“毛刺”,为观察数据局部细节,对260~500 s区间手机检测数据进行局部放大,如图3(b)所示。

5.2 数据重采样

手机标准采样频率100 Hz,对应的标准采样间隔为10 ms,但实际采样间隔的统计结果显示:智能手机采样时间间隔在区间 $[0, 11)$, $[11, 21)$ ms所占比例分别为51%和48%,其他占1%(剔除),如图4所示。对手机原生数据采取线性插值法进行数据重采样,可得均匀时间序列,如图5所示。

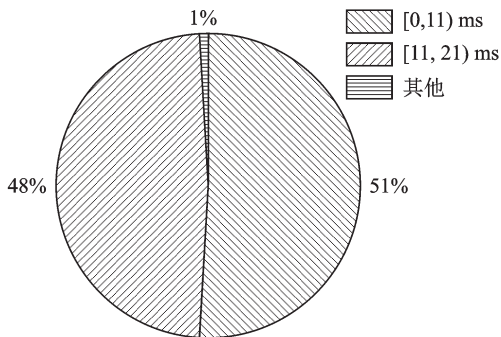


图4 手机采样间隔区间统计结果

Fig.4 Statistical results of phone sampling interval

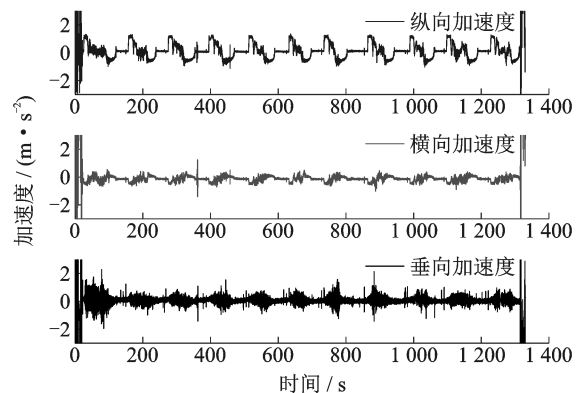


图5 手机检测数据(插值后)

Fig.5 Phone detection data (interpolation)

5.3 异常值识别

(1)根据上述异常值识别方法,首先基于 P 值筛选识别第一类异常值,设置了窗口长度 $N=100$,手机横向加速度阈值 $T_x=0.7 \text{ m/s}^2$,纵向加速度阈值 $T_y=1.25 \text{ m/s}^2$,垂向加速度阈值 $T_z=1.6 \text{ m/s}^2$, P_0 取为10%,则手机第一类异常值位置如图6中蓝色区域(幅值为1位置)。

(2)依据“ 3σ 准则”识别手机检测数据一阶变化率超限位置,手机第二类异常值分布位置如图6中红色部分(幅值为0.5位置)。

(3)针对手机加速度传感器信噪比低引起的随机误差,依据小波分析将手机检测数据分解为低频成份和细节成份,基于手机采样频率100 Hz和影响人体舒适度的频率范围0.5~80 Hz确定分解层数 $n=8$,选择尺度大小为2的Daubechies小波作为基小波,以手机垂向加速度为例,第7层、第8层细节成份幅值较小包含随机误差,其分解效果如图7所示。

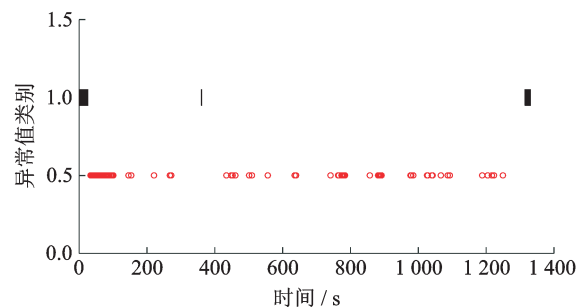


图6 手机检测数据异常值分布位置

Fig.6 Distribution of abnormal value of mobile phone detection data

5.4 异常值处理

(1)根据第4节异常值处理方法,因手机检测数据第一类异常值严重“失真”,采取区段删除将手机检测数据第一类异常值予以剔除,如图8所示。

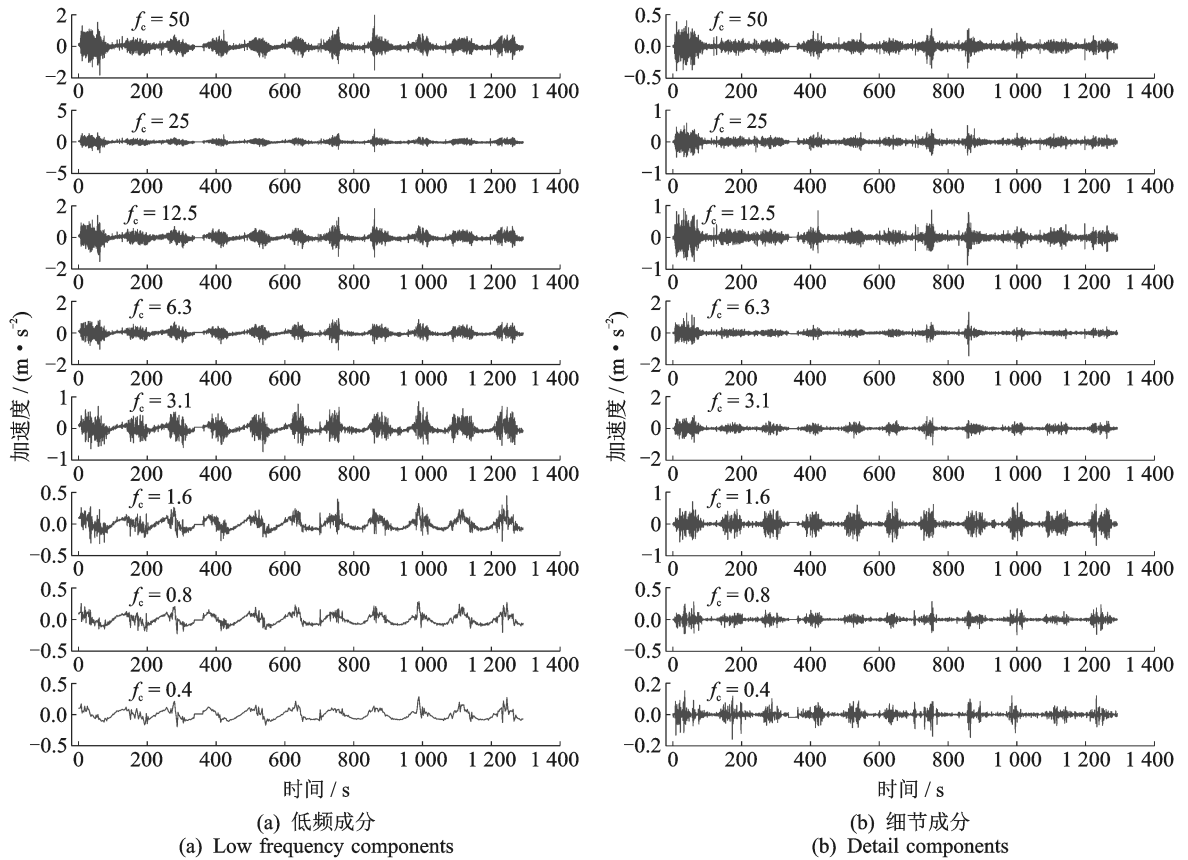


图7 手机检测数据小波分解

Fig.7 Wavelet decomposition of mobile phone detection data

(2)针对手机检测数据第二类异常值,其含有车体振动响应有效成份,将识别出的第二类异常值采用中值滤波依次处理,设置移动窗口 $N=5$,其滤除效果如图9所示。

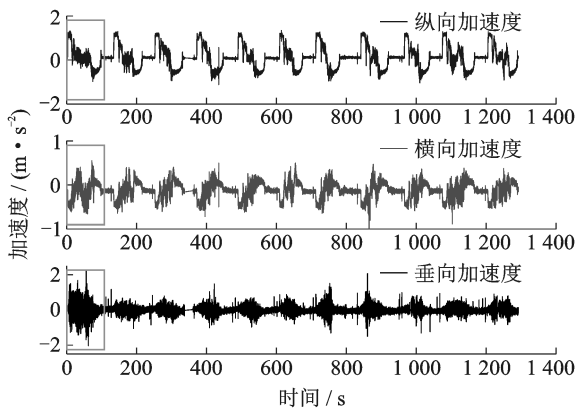


图8 剔除第一类异常值后手机检测数据

Fig.8 Detection data after eliminating the first kind of abnormal value

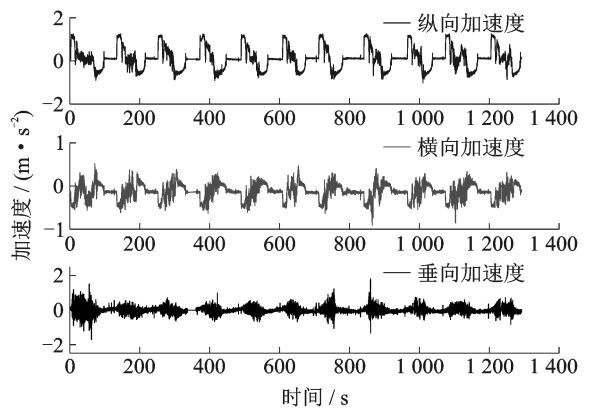


图9 剔除第二类异常值后手机检测数据

Fig.9 Detection data after eliminating the second kind of abnormal values

(3)根据第4节小波去噪原理,设置半软阈值将细节成份进行阈值量化,将随机误差成份滤除而保留手机检测数据有效成份,并将低频成份和细节成份逐层卷积求和实现重构,其处理效果如图10所示。

对比滤波前后手机检测数据局部细节,选取图8,图10中第一个运行区间[0,108]s手机检测数据,如图11所示,经中值滤波和小波滤波后,数据“毛刺”和随机误差被剔除。

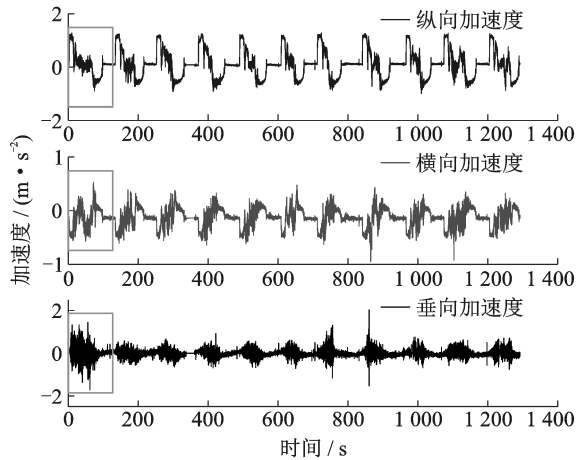


图10 小波分析滤波后手机检测数据

Fig.10 Mobile phone detection data after wavelet analysis filtering

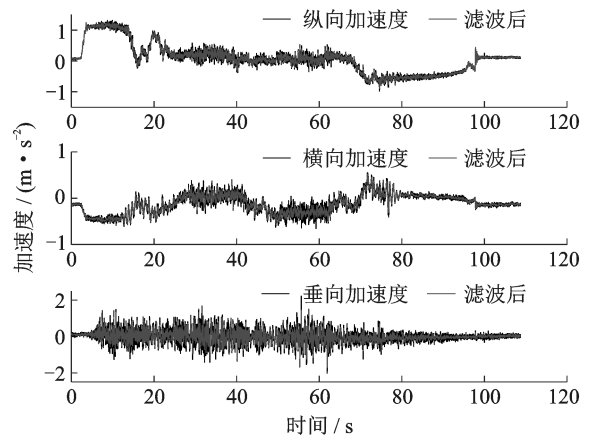


图11 小波分析滤波后局部手机数据对比

Fig.11 Comparison of local mobile phone data after wavelet analysis filtering

6 结束语

因手机检测数据异常值由多种原因引起,本文设置了相适应的异常值识别方法,可准确地识别出3种类型异常值的分布位置。同时为更好地滤除不同类型异常值,设置了相对应的剔除方法,为正确评价轨道平顺状态和车辆运行舒适性提供数据保障。本文研究可得以下结论:

(1)通过设置移动窗口 $N=100$,设置横向、纵向、垂向加速度阈值,当 $P > P_0$ 可定义为第一类异常值;将手机检测数据一阶变化率结合“ 3σ 准则”,若 $y_i > \mu + 3\sigma$ 或 $y_i < \mu - 3\sigma$ 可定义为第二类异常值;基于小波分析将手机检测数据分解为低频成份和细节成份,其中幅值较小的细节成份含有因传感器信噪比低引起的随机误差。

(2)提出了3种异常值适应性剔除方法:因手机受较大干扰引起第一类异常值,可予以剔除;因手机测试过程中偶然干扰引起第二类异常值,其含有车体振动有效成份,基于中值滤波滤除第二类异常值;因传感器信噪比低引起的随机误差可基于小波滤除。

(3)可实现高频次检测车体振动响应,弥补轨检车检测周期较长的不足。

参考文献:

- [1] 罗林. 轮轨系统轨道平顺状态的控制[M]. 北京: 中国铁道出版社, 2006.
- [2] Ping Wang, Yuan Wang, Li Wang, et al. Measurement of carbody vibration in urban rail transit using smartphones[C]// Transportation Research Board 96th Annual Meeting. [S.l.]: [s.n.], 2017: 15.
- [3] 毋红军, 刘章. 统计数据的异常值检验[J]. 华北水利水电大学学报(自然科学版), 2003(1): 72-75. Wu Hongjun, Liu Zhang. Abnormality test of statistical data [J]. Journal of North China University of Water Resources and Electric Power (Natural Science Edition), 2003(1): 72-75.
- [4] 郭志慧, 周傲英. 数据质量和数据清洗研究综述[J]. 软件学报, 2002, 13(11): 2076-2082.

- Guo Zhimao, Zhou Aoying. Research on data quality and data cleaning: A survey [J]. Journal of Software, 2002, 13(11): 2076-2082.
- [5] Chandola V, Banerjee A, Kumar V. Anomaly detection: A survey[J]. ACM Computing Surveys, 2009, 41(3): 1-58.
- [6] Chu X, Ilyas I F, Krishnan S, et al. Data cleaning: Overview and emerging challenges[C]//International Conference on Management of Data. [S.l.]: ACM, 2016: 2201-2206.
- [7] Garcia S, Luengo J, Herrera F. Data preprocessing in data mining[M]. Incorporated: Springer Publishing Company, 2016.
- [8] Chandrasekar P, Qian K. The impact of data preprocessing on the performance of a naive bayes classifier[C]//Computer Software and Applications Conference. [S.l.]: IEEE, 2016: 618-619.
- [9] 李再伟, 雷晓燕, 高亮. 轨道不平顺检测数据的预处理方法分析[J]. 铁道科学与工程学报, 2014(3): 43-47.
Li Zaiwei, Lei Xiaoyan, Gao Liang. Analysis of preprocessing method of track irregularity inspection data [J]. Journal of Railway Science and Engineering, 2014(3): 43-47.
- [10] 刘靖洁, 陈桂明, 刘晓芳, 等. FFT 和小波变换在信号降噪中的应用[J]. 数据采集与处理, 2009, 24(S1): 58-60.
Liu Qingjie, Chen Guiming, Liu Xiaofang, et al. Application of FFT and wavelet in signal denoising [J]. Journal of Data Acquisition and Processing, 2009, 24(S1): 58-60.
- [11] 王爱萍, 王惠南. 基于小波分析的 1/f 噪声降噪[J]. 数据采集与处理, 2006, 21(2): 217-221.
Wang Aiping, Wang Huinan. 1/f noise eliminating based on wavelet analysis [J]. Journal of Data Acquisition and Processing, 2006, 21(2): 217-221.
- [12] 吴小培, 柴晓冬. 一种中值滤波的快速算法[J]. 数据采集与处理, 1995(2): 151-155.
Wu Xiaopei, Chai Xiaodong. A fast algorithm of median filtering [J]. Journal of Data Acquisition and Processing, 1995(2): 151-155.
- [13] 张丽, 陈志强, 高文焕, 等. 均值加速的快速中值滤波算法[J]. 清华大学学报(自然科学版), 2004, 44(9): 1157-1159.
Zhang Li, Chen Zhiqiang, Gao Wenhuan, et al. Mean-based fast median filter [J]. Journal of Tsinghua University (Science and Technology), 2004, 44(9): 1157-1159.
- [14] Wei X, Jia L, Guo K, et al. On fault isolation for rail vehicle suspension systems[J]. Vehicle System Dynamics, 2014, 52(6): 847-873.
- [15] Lederman G, Chen S, Garrett J, et al. Track-monitoring from the dynamic response of an operational train [J]. Mechanical Systems and Signal Processing, 2017, 87: 1-16.
- [16] Wei X, Liu F, Jia L. Urban rail track condition monitoring based on in-service vehicle acceleration measurements[J]. Measurement, 2015, 80: 217-228.

作者简介:



从建力(1993-),男,硕士研究生,研究方向:轨道不平顺检测与评估,E-mail: cjl13618029520@163.com。



王源(1992-),男,博士研究生,研究方向:高速铁路轨道状态检测、评估、养护,信号处理,E-mail: 1220510620@qq.com。



杨翠平(1995-),女,硕士研究生,研究方向:轨道不平顺检测与评估。



王平(1969-),男,教授,研究方向:轨道不平顺及动力学、重载铁路钢轨波形磨耗,E-mail: wping@home.swjtu.edu.cn。



李成辉(1955-),男,教授,研究方向:轨道结构与轨道力学、道路工程力学行为分析与评价。

(编辑:夏道家)